1992

Integrated Stationary Time Series And Polyvariogram Methodology

Zhao-guo Chen

Integrated Stationary Time Series and Polyvariogram Methodology

by

Zhao-Guo Chen

Department of Statistical
and Actuarial Sciences

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
June 1992

ISBN 0-315-75359-5

Abstract

A time series, $\{Z(t)\}$, is called ***integrated stationary*** if there is a non-negative integer d such that $\{W(t) = \nabla^d Z(t)\}$ is second-order stationary, where ∇ is the backward differencing operator. A special case is the ARIMA(p, d, q) series.

By generalizing Cressie's (1988) results, the ***generalized*** and ***primary increment vectors of order b*** are defined. These vectors provide a more general way of transforming a homogeneous nonstationary series to stationarity than (repeated) simple differencing through the ∇^d operator. Resulting from the increment-vector methodology are several second-order moment characterizations of an integrated stationary series: the ***variogram,*** the ***generalized covariance function*** and two types of ***polyvariogram.*** The interrelationships and properties of these characterizations are investigated.

The increment-vector methodology also leads to a representation of $\{Z(t)\}$ which reflects the intrinsic stochastic nature of the series and is proved to agree with a special representation defined by Matheron (1973) for spatial processes. From our representation, a theorem for a decomposition of the series is deduced. Using the representation and decomposition theorem we can analyse what happens when overdifferencing takes place and what the divergence rate of the series is (a law of the iterated logarithm is proved).

For two types of polyvariogram of order b, where integer $b \geq \max\{0, d-1\}$, the general formulae of the asymptotes are obtained, which show a positive slope when $b = d-1$ and zero slope when $b > d-1$. This is the key feature of the polyvariograms which is exploited in the graphical identification of d. The asymptotic distribution and the almost sure convergence rate of the sample polyvariograms are obtained for various b, d and $\{W(t)\}$. Based on these results, we propose some statistical testing procedures for determining d given an integrated white noise or an integrated ARMA series.

Acknowledgements

I do not have adequate words to express my appreciation and gratitude to Dr. O.D. Anderson who invited me to visit this department and later on became my supervisor. Thanks to his strong, thoughtful and enduring support, I have been extremely fortunate for the past two years to spend the most successful period of my life in this university. The countless stimulating discussions, the harmonious and enjoyable cooperation, and his heart-warming encouragement will live in my memory and influence my future.

I am deeply indebted to Dr. I.A. McLeod. Due to his kind concern and recommendation, the winning of my various awards became reality. He bent over backwards to help me in many respects and even came to my home to install software on my PC which has been very helpful for producing this thesis. I have greatly benefited from many useful conversations with him.

I very much appreciate the extremely valuable advice, encouragement and help from Dr. I.B. MacNeill.

I am grateful to Dr. M.S. Haq, Dr. M.M. Ali, Dr. R.J. Kulperger, Dr. D.R. Bellhouse and many other faculty members in the Department for their concern, support and help. I also thank Mr. L.M. Kwarciak and Mrs. M.I. Lavdas for their computer services and LaTeX consulting. I am appreciative of much help from Corinne Harrison, Alicia Pleasence and Lisa Bouchard.

I warmly acknowledge the Faculty of Graduate Studies for their generous financial support and kindness during my stay.

I am very obliged to Dr. N. Cressie. Due to his invitation, I had an opportunity to visit Iowa State University in early 1989 and became greatly interested in his ingenious work, which inspired the content of this thesis.

I thank my family for all their support and understanding during my busiest years; especially my wife, Zhong-Qin Wu, who typed and retyped the thesis with fortitude.

TABLE OF CONTENTS

Chapter 1

Transformation from Nonstationarity

1.1 Introduction

1.1.1 Integrated stationary time series

Let $\{Z(t)\} = \{Z(t) : t = 0, 1, \ldots\}$ denote a time series. Denote the backshift operator by B and the backward difference operator by $\nabla = 1 - B$. Suppose $d \geq 0$ is an integer and

$$\nabla^d Z(t) = W(t), \qquad t = d, d+1, \ldots, \tag{1.1.1}$$

where $\{W(t)\} = \{W(t) : t = \ldots, -1, 0, 1, \ldots\}$.

When $\{W(t)\}$ is a stationary and invertible ARMA (p, q) series, the $W(t)$ satisfy

$$\phi(B)W(t) = \theta(B)A(t); \tag{1.1.2}$$

where the $A(t)$ are i.i.d. or more general white noise shocks, with variance σ^2, and $\phi(\zeta)$ and $\theta(\zeta)$ are relatively prime polynomials such that

$$\begin{aligned} \phi(\zeta) &\equiv 1 - \phi_1\zeta - \cdots - \phi_p\zeta^p \neq 0 \\ \theta(\zeta) &\equiv 1 - \theta_1\zeta - \cdots - \theta_q\zeta^q \neq 0 \end{aligned} \qquad |\zeta| \leq 1. \tag{1.1.3}$$

Then $\{Z(t)\}$ becomes an invertible ARIMA(p, d, q) series whose terms satisfy

$$\phi(B)\nabla^d Z(t) = \theta(B)A(t), \quad t \geq p + d, \tag{1.1.4}$$

which is nonstationary when $d > 0$. The acronym ARIMA signifies an *integrated* ARMA series.

However, we consider a much more general situation : $\{W(t)\}$ is second-order stationary, that is

$$\begin{aligned} &E\{W(t)\} = 0 \\ &E\{W(t)W(t \pm h)\} = \gamma(h), \quad h = 0, 1, \ldots, \end{aligned} \tag{1.1.5}$$

where $\gamma(h)$ is the autocovariance function (ACVF) and there is no loss in generality from imposing the condition $E\{W(t)\} = 0$. When (1.1.5) holds, we call the $\{Z(t)\}$ in (1.1.1) an *integrated stationary time series*, as $Z(t)$ may then be obtained by integrating the stationary series, $\{W(t)\}$. If, further, $\{W(t)\}$ does not result from overdifferencing [i.e., $\{\nabla^{d-1}Z(t)\}$ is not second-order stationary], then we call $\{W(t)\}$ the *hub series* of $\{Z(t)\}$.

1.1.2 Some literature concerning the identification of d

Model (1.1.1) has been of wide-spread interest over the past two decades, especially due to the popularization of its special case, the ARIMA (p, d, q) model. There are many papers treating model (1.1.1) under various assumptions, and a major research topic is the identification of d. The approaches in the literature for identifying d may be classified into the following categories.

The AR-fitting approach

Assume the $Z(t)$ satisfy

$$\psi(B)Z(t) = X(t); \tag{1.1.6}$$

where $\psi(\zeta) \equiv 1 - \psi_1\zeta - \cdots - \psi_{p+d}\zeta^{p+d} \neq 0$ when $|\ \zeta\ | < 1$, and $\{X(t)\}$ may be more general than merely an i.i.d. series. [In the ARIMA case, $\psi(B) = \phi(B)\nabla^d$ and $\{X(t)\}$ is MA(q)]. Let $\boldsymbol{\psi} = (\psi_1 \cdots \psi_{p+d})$ and, to obtain $\hat{\boldsymbol{\psi}}$, use ordinary least squares or some other estimation procedure [that of Tsay and Tiao 1984, say, for ARIMA (p, d, q)]. Next, if one can either prove that $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \to 0$ at a particular rate, in some probabilistic sense (see, say, Tiao and Tsay 1983, Huang 1988), or obtain the asymptotic distribution of $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}$, then inferential methodology may be developed to judge whether $\psi(B)$ contains the factor ∇^d for some d [but not the factor ∇^{d+1}].

For deriving the asymptotic distribution of $\hat{\psi} - \psi$, the earliest contribution is White (1958) and more influential recent work is due to Dickey and Fuller (say, 1979). But these approaches only consider $\psi(B) = 1 - \psi_1 B$ and the $X(t)$ i.i.d.. [White also assumed the $X(t)$ to be Gaussian.] Dickey and Fuller test the null hypothesis, $\psi_1 = 1$, using the distribution that they derive — to give a so-called "unit-root test". There are many subsequent extensions to this approach. For instance: Said and Dickey (1985) deal with ARIMA $(p, 1, q)$; Phillips and Perron (1988) again consider $\psi(B) = 1 - \psi_1 B$, but assume $\{X(t)\}$ is subject to strong mixing; Chan and Wei (1988) examine ARIMA $(p, d, 0)$ and obtain the specific asymptotic distribution when $p = 0$.

Approaches based on AIC-like criteria

These transform the $Z(t)$ data to $\nabla^b Z(t)$ $(t \geq b)$ for various b. Put $S(b) = \sum_t \{\nabla^b Z(t)\}^2$. Theoretically, this $S(b)$ decreases dramatically, with b increasing up to and including d, but then changes only relatively slowly as b increases further. Consider modifying $S(b)$, by adding a suitable increasing function of b (which penalises overdifferencing), so that the modified quantity should take an unequivocal minimum value at $b = d$. Huang (1988) worked in this way for ARIMA (p, d, q) series and proved his procedure was strongly consistent. In fact, earlier than that, Yajima (1985a) used long AR approximations to $\nabla^b Z(t)$, handled the squared sum of residuals in a similar manner, and proved the consistency of his method.

The sample autocorrelation function approach

Given any series of length n, say $Y(1), \ldots, Y(n)$, the sample autocovariance and autocorrelation functions (abbreviated to SACVF and SACF, respectively) can always be defined, for $h = 0, 1, \ldots, k$ $(k \leq n - 1)$, as:

$$\hat{\gamma}^{(Y)}(h) = \frac{1}{n} \sum_{t=1}^{n-h} \{Y(t) - \overline{Y}\}\{Y(t+h) - \overline{Y}\}, \qquad (1.1.7)$$

where $\overline{Y}$ is the sample mean, and

$$\hat{\rho}^{(Y)}(h) = \hat{\gamma}^{(Y)}(h)/\hat{\gamma}^{(Y)}(0). \qquad (1.1.8)$$

Observe the SACF of $\{Y(t) = \nabla^b Z(t+b)\}$ for some $b \geq 0$. Box and Jenkins (1976) argued that a tendency for this SACF to decay only slowly from a lag-zero value of 1 suggests the nonstationarity of $\{\nabla^b Z(t)\}$ but leaves the question of the nonstationarity or stationarity of $\{\nabla^{b+1} Z(t)\}$ open. This offers an identification criterion for d.

Many authors have tried to tie down this feature more precisely. For a once-integrated MA(1), Wichern (1973) had obtained $E\{\hat{\rho}^{(Z)}(h)\}$ approximately and showed that, when $\theta_1 > 0$ (but θ_1 was not too near to 1 or 0, given n), the slow decay in the SACF could be from a positive value considerably less than 1. Anderson (1979) generalized Wichern's work to ARIMA$(p,1,q)$ and obtained corresponding formulae for ARMA(p,q) and ARIMA(p,d,q), $d \geq 2$. This demonstrated that: Wichern-like behaviour of a slow (ragged) linear decay, emanating from a positive value less than one, indicated $d = 1$ or 0; a very smooth decay from one with a slope close to $-3/n$ indicated $d \geq 2$; a likely less smooth decay, again from one, with slope of roughly $-5/n$ suggested $d = 1$; while a (more ragged and) considerably faster decline from one suggested $d = 0$ with a zero of the stationary AR operator approaching one. The problem with this contribution is that, although it sharpened the Box-Jenkins tool for identifying d, it did not provide a formal mechanism for testing an identified d.

Still with the theme of a slow linear decay in the SACF, Hasza (1980) showed that, for $d = 1$, $n\{1 - \hat{\rho}^{(Z)}(h)\} \xrightarrow{L} R(h)$, a random variable which increases approximately linearly with increasing h; while, under more general model assumptions, Yajima (1985b) proved that, for $d \geq 1$, $n\{1 - \hat{\rho}^{(Z)}(h)\} \xrightarrow{L} hZ_5$, where Z_5 is a random variable. Anderson (1990) also discussed some features of cross-over points which characterize (and distinguish between) $d = 0$, $d = 1$ and $d \geq 2$ (but again with the emphasis on identification rather than testing).

1.1.3 The variogram and other second-order moment characterizations

The autocorrelation function (ACF) [or autocovariance function (ACVF)] plays a central role in describing the statistical properties of second-order stationary series;

and, based on it, many inferential and modelling procedures have been developed. For stationary series, the SACF of form (1.1.8) [or SACVF, (1.1.7)] has a variety of nice properties of convergence to its theoretical ACF (or ACVF) under various assumptions. However, the ACF does not exist for nonstationary series; and although, as a practical tool, the SACF sometimes works in identifying d, dissatisfaction with the "slow-decay" criterion of Box and Jenkins has often been reported [e.g., Anderson 1985]. In practice, for $d = 0$ and 1, it is often ambiguous as to whether the SACF dies away slowly or quickly.

The approach of AR-fitting and those based on AIC-like criteria are neat and can be developed rigorously. But, for each, there only then results a single quantity (or just a few parameter estimates) to summarize the information, contained in the data, which is then frequently inadequately captured. For instance, if the p-value obtained from a unit-root test is 0.06, how much confidence do we have in a decision of accepting the null hypothesis of $\psi_1 = 1$? In such a situation, the pragmatist would expect additional information in a visual format, as the SACF provides in the stationary case. Moreover, the current results achieved with these various approaches generally require extra, sometimes quite stringent, conditions rather than just the simple second-order stationarity of $\{W(t)\}$, or to be restricted to the case of $d = 1$ (for instance, Phillips and Perron 1988, although second-order stationarity is replaced there by other weak conditions).

There is, however, a suitable second-order moment statistical description for series, $\{Z(t)\}$, that are either stationary or nonstationary (of the once-integrated stationary type), viz.:

$$Y_0(h,t) = Z(h+t) - Z(t), \quad t = 0,1,\ldots, \quad h = 0,1,\ldots, \tag{1.1.9}$$

$$v_0(h) = \text{Var}\{Y_0(h,t)\} = E\{Y_0^2(h,t)\}, \quad h = 0,1,\ldots \,. \tag{1.1.10}$$

The quantity, $v_0(h)$, has been called a *structure function* in physics (Kolmogorov 1941), a *mean squared difference* in the time series literature (Jowett 1952); and Matheron (1963) called it a *variogram* in geostatistics. (Here we add the subscript 0 to Y and v to indicate the *order* of these quantities, as later we will be introducing a general order b.)

Let us see how the variogram works in distinguishing $d = 1$ from $d = 0$ for model (1.1.1). When $d = 0$, $v_0(h)$ becomes

$$v_{00}(h) = E[\{W(h+t) - W(t)\}^2] = 2\{\gamma(0) - \gamma(h)\}, \quad h = 1, 2, \ldots, \tag{1.1.11}$$

where $\gamma(h)$ is defined in (1.1.5) and we add the second subscript 0 to v to indicate the special expression for $v_0(h)$ in the case of $d = 0$. When $\{W(t)\}$ is regular (has no purely deterministic part in its Wold decomposition), then $\gamma(h) \to 0$ as $h \to \infty$, and hence $v_{00}(h)$ has a horizontal line [with intercept $2\gamma(0)$] as its asymptote.

When $d = 1$, let us just consider the simplest case for now, where $\{W(t)\}$ is white noise. [For the general case, see (2.2.32) below.] Then $\{Z(t)\}$ is a random walk, and $v_0(h)$ becomes

$$v_{01}(h) = E[\{\sum_{j=t+1}^{t+h} A(j)\}^2] = h\sigma^2, \quad h = 1, 2, \ldots; \tag{1.1.12}$$

i.e., $v_{01}(h)$ is a straight line with positive slope, σ^2.

Thus, the variogram is well-defined for model (1.1.1), when $d = 0$ or $d = 1$, and has strikingly distinct visual features in these two cases. So it offers another means of identifying d — and, by transmitting more information than the approaches surveyed in subsection 1.1.2, the variogram promises to be a more effective tool for determining d.

Let $\boldsymbol{\pi}_0^h$ denote the $(h+1)$-vector $(1\ 0 \cdots 0\ -1)'$ and $\mathbf{Z}_h(t) = (Z(h+t)\ Z(h-1+t)\ \cdots Z(t))'$. Then, from (1.1.9), we may write

$$Y_0(h,t) = \mathbf{Z}_h'(t)\boldsymbol{\pi}_0^h, \quad t = 0, 1, \ldots . \tag{1.1.13}$$

The family of vectors, $\{\boldsymbol{\pi}_0^h : h = 1, 2, \ldots\}$, transforms any $\{Z(t)\}$ to a family of series $[\{Y_0(h,t)\} : h = 1, 2, \ldots]$; from which the variogram, $v_0(h)$, is defined by (1.1.10). However, (1.1.10) is only defined for $d = 0$ and $d = 1$. For higher d, $\{Y_0(h,t) : t = 0, 1, \ldots\}$ is no longer second-order stationary and $\text{Var}\{Y_0(h,t)\}$ depends on t [which renders (1.1.10) meaningless]. So, when $d > 1$, we need to find further families of vectors that play roles similar to $\{\boldsymbol{\pi}_0^h : h = 1, 2, \ldots\}$.

Although Matheron (1973) did provide some general theory for dealing with nonstationarity in spatial data, time series researchers had largely ignored this very promising advance until Cressie (1988) ingeniously introduced his ***primary increment vectors*** (PIV) of order b, $b \in \{0, 1, 2\}$, and then defined his ***semivariogram, linvariogram*** and ***quadvariogram*** which indicated a way of dealing with (1.1.1) for lower d. In fact, these quantities defined by Cressie are special cases of Matheron's ***general covariance*** (GC) — called the ***general covariance function*** [GCF, see (1.4.1) below] in time series analysis. The GCF provides a second-order moment characterization for integrated stationary time series.

There is another quite well-known second-order moment characterization which is due to Yaglom (see, say, 1962, p 88; 1987, p 427), called the ***structure function.*** It was defined for continuous ***processes with stationary increments of order d*** (as Yaglom called them). Observing such a process at equispaced time points, we obtain a time series satisfying model (1.1.1) (but Yaglom allows $t \rightarrow -\infty$); and then it can be shown that Yaglom's structure function may be obtained from the GCF [by specially choosing $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ in (1.4.1)].

In this thesis we will introduce a second-order moment characterization, $v_b(h)$, called the ***variogram of order b***, which can be obtained from the GCF [by choosing $\boldsymbol{\lambda} = \boldsymbol{\mu} = \boldsymbol{\pi}_b^h$ in (1.4.1)]. So, among all these characterizations, the GCF is the most general; but, except for the case of order $b = 0$, it is not unique. One of the uses of $v_b(h)$ is to retrieve the GCF (to within a certain polynomial). For $b = 0$, the structure function and our variogram are the same as the two quantities defined by, respectively, Kolmogorov (1941) and Matheron (1963).

1.1.4 Outline of thesis

This chapter develops a general theory of increment vectors — a bridge between nonstationarity and stationarity, and a tool for characterizing and classifying nonstationarity. Section 1.2 just gives a little further notation and some results which relate to it. In Section 1.3, we introduce the GCF, redefine a ***general increment vector*** (GIV) more conveniently and give a decomposition for it. This decomposition leads to a simplification of Cressie's definition of an I_d-process (or I_d-series, in our terminology) — which permits us to directly interpret (1.1.1) as such a series. In Section 1.4, for $b = 0$, 1 and 2, we define a variogram of order b using Cressie's PIV, and show how to use the PIV to connect this variogram with the GCF by a difference equation — which in fact indicates a generalization of what Cressie did. Motivated by this connection, Section 1.5 introduces the PIV of order b for any non-negative integer b (which, for $b \in \{0, 1, 2\}$, agrees with Cressie's definitions). We are then able to define the variogram of general order b. Section 1.6 suggests improvements to the presentation of the increment-vector methodology, if one is prepared to ignore certain precedents set in the literature.

Chapter 2 studies the properties of some second-order moment statistical descriptions which characterize integrated stationary time series: the variogram and the GCF, and the subsequently defined quantities — two types of *polyvariogram*. Section 2.1 discusses the solution of the difference equation mentioned in Section 1.4, so that we may express the GCF through the variogram. This provides a theoretical underpinning for the estimation of the GCF, since the variogram can be estimated directly from data $Z(0), \ldots, Z(n)$. Section 2.2 derives formulae for the variogram via the ACVF of the hub series; and, in doing so, some other useful subsets of the GIV, called *integrated* PIV, are introduced. As the GCF is not unique, we choose a canonical one (called the *primary* GCF), and Section 2.3 obtains formulae for it via the ACVF of the hub series. From these results, we may achieve the asymptotes of two types of polyvariogram (defined in terms of, respectively, the variogram and the primary GCF), which play a key role in the graphical identification of d. Section 2.4 is dedicated to this purpose.

Chapter 3 investigates the structure of realizations of integrated stationary time series. Section 3.1 defines an informative representation and proves a basic decomposition theorem. Section 3.2 demonstrates some applications of this theory, including: divergency analysis, insight into overdifferencing, and forecasting. Using the decomposition theorem, Section 3.3 obtains a law of the iterated logarithm which sharpens the result of Lai and Wei (1983) and precisely describes the divergency behaviour of

I_d-series realizations.

Chapter 4 develops inferential procedures for determining d, the correct degree of differencing, mainly using *scaled sample polyvariograms* (SSPV). Section 4.1 defines the statistics used for this purpose, discusses their unbiasedness and proposes a graphical procedure based on that of Cressie (1988). Using the formulae which are established in Sections 2.2 and 2.3, Section 4.2 investigates the asymptotic properties of the two types of sample polyvariogram under the assumption that the hub series are general linear series. Section 4.3 sets up the relevant hypotheses and shows that the SSPV can only be used for testing the degree of differencing in the integrated white noise case. A central limit theorem and a law of the iterated logarithm are proved for the SSPV of integrated white noise, and some procedures for testing based on these results are proposed. Section 4.4 considers how to solve a problem of testing d for an ARIMA series by solving the same problem for an approximate integrated white noise, obtained by transfomation, which then allows the SSPV to be used as test statistics — and the corresponding asymptotic results are derived for this transformed model. Section 4.5 provides an appendix to Section 4.4.

1.2 Some Further Notation and Related Results

Throughout this thesis, we keep t , the standard convention that a summation is zero when its lower limit exceeds its upper one. We also denote the full set of integers by $\mathcal{Z}$, the set of non-negative integers by $\mathcal{N}$, the set of strictly positive integers by $\mathcal{Z}^+$, the set of all real values by $\mathcal{R}$, and let $\mathcal{R}^n$ be the set of all n-vectors whose components are elements of $\mathcal{R}$.

For the difference equation results which we use, see say Hildebrand (1968) or Lakshmikantham and Trigiante (1988).

1.2.1 Differencing and summation operators

Corresponding to ∇, we also denote the forward difference operator by $\Delta = B^{-1}\nabla$; such that, for any function [$f(t)$ say], $\Delta f(t) = f(t+1) - f(t)$. Then

$$\nabla^b f(t) = (1-B)^b f(t) = \sum_{i=0}^{b}(-1)^i \binom{b}{i} f(t-i), \quad b \in \mathcal{N}, \tag{1.2.1}$$

and

$$\Delta^b f(t) = (B^{-1}-1)^b f(t) = (-1)^b \sum_{i=0}^{b}(-1)^i \binom{b}{i} f(t+i), \quad b \in \mathcal{N}. \tag{1.2.2}$$

A related summation operator, Σ, can be defined by

$$\Sigma g(t) = \sum_{i=r}^{t-1} g(i) \quad (\forall t \geq r), \tag{1.2.3}$$

where the r, here and elsewhere in this section, is some fixed integer from $\mathcal{Z}$. [When $g(t)$ is $Z(t)$, a convenient choice would be $r = 0$. However, in this paper, we mainly use the operator with functions of h which are only defined for $h \in \mathcal{Z}^+$. It is then appropriate to choose $r = 1$.] Note that, according to (1.2.3), $\Sigma g(r) = 0$.

1.2.2 Several lemmas

Lemma 1.2.1 *For $b \in \mathcal{Z}^+$ and all $t \geq r$,*

$$\Sigma^b g(t) = \frac{1}{(b-1)!}\sum_{i=r}^{t-b}(t-1-i)^{(b-1)} g(i). \tag{1.2.4}$$

The bracketed superscript, here and later on, indicates a "factorial power"; that is

$$s^{(m)} = \begin{cases} 1, & m = 0, \\ s(s-1)\cdots(s-m+1), & m \geq 1. \end{cases} \tag{1.2.5}$$

This time, we note that $\Sigma^b g(t) = 0$ for those $t \in \{r, r+1, \ldots, r+b-1\}$.

Proof We prove the result by induction. Definition (1.2.3) provides the base case, with $b = 1$. Now, whenever a case with $b \geq 1$ is true, we have [again using (1.2.3)]:

$$\begin{aligned} \Sigma^{b+1} g(t) &= \Sigma\{\frac{1}{(b-1)!}\sum_{i=r}^{t-b}(t-1-i)^{(b-1)} g(i)\} \\ &= \frac{1}{(b-1)!}\sum_{j=r}^{t-1}\{\sum_{i=r}^{j-b}(j-1-i)^{(b-1)} g(i)\} \end{aligned}$$

$$= \frac{1}{(b-1)!}\sum_{i=r}^{t-1-b}\{\sum_{j=i+b}^{t-1}(j-1-i)^{(b-1)}\}g(i)$$

$$= \frac{1}{(b-1)!}\sum_{i=r}^{t-1-b}\sum_{j=i+b}^{t-1}\frac{1}{b}\{(j-1)^{(b)}-(j-1-i)^{(b)}\}g(i)$$

$$= \frac{1}{b!}\sum_{i=r}^{t-b-1}(t-1-i)^{(b)}g(i).$$

So the inductive step is proven. □

Lemma 1.2.2 *For $b \in \mathcal{Z}^+$, the difference equation*

$$\Delta^b f(t) = g(t) \tag{1.2.6}$$

has

$$f(t) = \Sigma^b g(t) \tag{1.2.7}$$

as a particular solution, for all $t \geq r$.

Proof First note that the operator Σ is a right inverse of Δ, for

$$\Delta\Sigma g(t) = \Delta\{\sum_{i=r}^{t-1} g(i)\} = (\sum_{i=r}^{t} - \sum_{i=r}^{t-1})g(\cdot) = g(t).$$

Try $f(t) = \Sigma^b g(t)$ as a solution in (1.2.6). Then:

$$\text{LHS} = \Delta^b\Sigma^b g(t) = \Delta^{b-1}(\Delta\Sigma)\Sigma^{b-1}g(t) = \Delta^{b-1}\Sigma^{b-1}g(t) = \cdots = g(t) = \text{RHS}.$$

So, $f(t) = \Sigma^b g(t)$ is a particular solution of (1.2.6). □

It is well-known that repeatedly simple-differencing a proper polynomial of degree $j \geq 0$, $f_j(t)$ say, a non-negative integral number of times, b, will reduce the polynomial to a related proper one of degree $j - b$ when $j \geq b$, and annihilate the polynomial when and only when $j < b$. Thus, using (1.2.1) and (1.2.2), when j and $b \in \mathcal{N}$:

$$\sum_{i=0}^{b}(-1)^i\binom{b}{i}f_j(t \pm i)\begin{cases}\text{is a proper polynomial} & \\ \text{of degree } j-b & \text{if} \quad j \geq b \\ \equiv 0 & \text{iff} \quad j < b;\end{cases} \tag{1.2.8}$$

and, in particular, on choosing $f_j(t) = t^j$ and putting $t = 0$, (1.2.8) yields

$$\sum_{i=0}^{b}(-1)^i \binom{b}{i} i^j \begin{cases} \neq 0 & \text{if} \quad j = b \\ = 0 & \text{if} \quad j < b. \end{cases} \tag{1.2.9}$$

The following lemma is easy to check (see Hildebrand 1968, p 19).

Lemma 1.2.3 *For any $k \in \mathcal{Z}^+$,*

$$\Delta t^{(k)} = kt^{(k-1)}, \quad t \in \mathcal{R}; \tag{1.2.10}$$

$$\nabla t^{(k)} = k(t-1)^{(k-1)}, \quad t \in \mathcal{R}; \tag{1.2.11}$$

and, for any $k \in \mathcal{N}$,

$$\Sigma t^{(k)} = \sum_{i=1}^{t-1} i^{(k)} = t^{(k+1)}/(k+1), \quad t \in \mathcal{Z}^+. \tag{1.2.12}$$

Lemma 1.2.4 *For any $k \in \mathcal{N}$ and $\mu \geq 0$, as $t \to \infty$,*

$$t^{(k)} = t^k + O(t^{k-1}); \tag{1.2.13}$$

$$\Sigma t^\mu = \sum_{i=1}^{t-1} i^\mu = O(t^{\mu+1}). \tag{1.2.14}$$

Proof (1.2.13) follows from the definition of $t^{(k)}$. (1.2.14) is evident on noting that

$$\frac{(t-1)^{\mu+1}}{\mu+1} = \int_0^{t-1} s^\mu ds \leq \sum_{s=1}^{t-1} s^\mu \leq \int_1^t s^\mu ds = \frac{t^{\mu+1}-1}{\mu+1}. \tag{1.2.15}$$

Lemma 1.2.5 *For $b \in \mathcal{Z}^+$,*

$$\Delta^b\{u(t)v(t)\} = \sum_{i=0}^{b} \binom{b}{i} \{\Delta^i u(t)\}\Delta^{b-i} v(t+i). \tag{1.2.16}$$

Proof It is easy to check that

$$\Delta\{u(t)v(t)\} = u(t)\Delta v(t) + \{\Delta u(t)\}v(t+1). \tag{1.2.17}$$

Then (1.2.16) follows by induction. □

1.2.3 The assign notation

Finally, we introduce the *assign* operator, A, which transforms a general (possibly degenerate) m-degree polynomial operator into a corresponding $(m+1)$-column vector. If $f(B) = \sum_{i=0}^{m} f_i B^i$, then we write

$$A\{f(B)\} = (f_0 \quad f_1 \quad \cdots \quad f_m)' = \mathbf{f}, \tag{1.2.18}$$

say; where the inverse operator A^{-1} (which undoes A) is equivalent to premultiplying by the $(m+1)$-row vector, $(1 \; B \; \cdots \; B^m)$, so that

$$A^{-1}(\mathbf{f}) = f(B). \tag{1.2.19}$$

1.3 Fundamental and Generalized Increment Vectors and I_d and $\tilde{I}_d$ Series

1.3.1 Matheron's characterization for nonstationary spatial processes

Matheron (1973) proposed a fundamental way of using second moments to characterize a certain kind of nonstationary spatial process, $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{R}^n\}$. For $b \in \mathcal{N}$, Matheron introduced a set of measures defined on $\mathcal{R}^n$, $\Lambda_b = \{\lambda\}$ (with a finite support for each λ), all of which annihilate any polynomial of degree up to b and are closed under translation. That is: $\lambda(d\mathbf{x}) = 0$, for all but a finite number of $\mathbf{x} = (x_1, \ldots, x_n)'$;

$$\int x_1^{i_1} \cdots x_n^{i_n} \; \lambda(d\mathbf{x}) = 0, \qquad \text{if} \quad i_1 + \cdots + i_n \leq b; \tag{1.3.1}$$

and $T^{\mathbf{y}}\lambda(d\mathbf{x}) \overset{\text{def}}{=} \lambda\{d(\mathbf{x}+\mathbf{y})\} \in \Lambda_b$, for any $\mathbf{y} \in \mathcal{R}^n$. (The notation $\overset{\text{def}}{=}$ denotes "is defined by".)

Suppose $Z(\mathbf{x})$ is a random function which maps $\mathbf{x}$ from $\mathcal{R}^n$ onto $L^2(\Omega, \mathcal{A}, P)$, and define $Z^*(\lambda) = \int Z(\mathbf{x})\lambda(d\mathbf{x})$. Then Matheron termed $Z(\mathbf{x})$ an intrinsic random function (IRF) of order b, if $Z^*(\lambda)$ were stationary that is: if, for any $\lambda \in \Lambda_b$, the group of shift operators, $U^{\mathbf{y}}$ ($\mathbf{y} \in \mathcal{R}^n$) such that $U^{\mathbf{y}} Z^*(\lambda) = Z^*(T^{\mathbf{y}}\lambda)$, is unitary. Thus, for any $\lambda, \mu \in \Lambda_b$,

$$\text{Cov}\{Z^*(T^{\mathbf{y}}\lambda), Z^*(T^{\mathbf{y}}\mu)\} = \text{Cov}\{U^{\mathbf{y}} Z^*(\lambda), U^{\mathbf{y}} Z^*(\mu)\} = \text{Cov}\{Z^*(\lambda), Z^*(\mu)\}. \tag{1.3.2}$$

Then, if a function $\kappa(\mathbf{h})$ on $\mathcal{R}^n$, with the "symmetric" property $\kappa(\mathbf{h}) \equiv \kappa(-\mathbf{h})$, satisfies

$$\mathrm{Cov}\{Z^*(\lambda), Z^*(\mu)\} = \int \lambda(d\mathbf{x})\kappa(\mathbf{x}-\mathbf{y})\mu(d\mathbf{y}) \qquad (\forall\lambda, \mu \in \Lambda_b), \tag{1.3.3}$$

this $\kappa(\mathbf{h})$ is called a generalised covariance function (GCF) of $\{Z(\mathbf{x})\}$. Thus, $\kappa(\mathbf{h})$ describes the second-order moments of $Z(\mathbf{x})$; and, strictly speaking, should be referred to as the GCF of order b, and denoted by, say, $\kappa_b(\mathbf{h})$.

1.3.2 The integrated stationary time series specialisation

For a time series, $\{Z(t) : t = 0, 1, \ldots\}$, we have $n = 1$. So, then, the GCF reduces to $\kappa_b(h)$, an even function due to the "symmetric" property, and we can restrict Λ_b to measures which, at most, have masses only at points in $\mathcal{N}$ — and the translation (shift) distances are also integral. For example, if a λ just has a mass of 1 at h ($h \geq 1$) and another of -1 at 0, then $\lambda \in \Lambda_0$, as λ annihilates any constant c (i.e. any polynomial of degree 0) since $\int c\lambda(dt) = c - c = 0$. So, if $\nabla Z(t) = W(t)$, $Z^*(\lambda) = Z(h) - Z(0) = \sum_{t=1}^{h} W(t)$; and, after translation, $T^k\lambda$ has just one mass of 1 at $h + k$ and another of -1 at k, and $Z^*(T^k\lambda) = Z(h+k) - Z(k) = \sum_{t=1+k}^{h+k} W(t)$. Again, if $U^kW(t) = W(t+k)$, then clearly $U^kZ^*(\lambda) = Z^*(T^k\lambda)$ — but note that we cannot immediately assert that $Z(t)$ is an IRF of order 0, as we have not yet verified that $U^kZ^*(\lambda) = Z^*(T^k\lambda)$ for all $\lambda \in \Lambda_0$. (This point will be taken care of by Theorem 1.3.1, below.)

The example indicates that, for time series, corresponding to a measure, $\lambda \in \Lambda_b$, there is a vector of masses; and, in the sequel, we will denote this vector by $\boldsymbol{\lambda}$. To avoid extra notation, Λ_b will also be used to indicate the corresponding set of vectors (the context each time making it clear whether a set of measures or vectors is implied). Cressie (1988) referred to such a vector as a generalised increment-vector (GIV) of order b. Corresponding to the example, $(1 \quad 0 \cdots 0 \quad -1)'$ (with a string of $h-1$ zero elements) provides an example of a GIV, with $b = 0$; and reference to a GIV will also eliminate the need for a measure-theoretic framework in much of what follows.

1.3.3 The FIV and GIV

Suppose $\boldsymbol{\lambda}$ is a column vector of finite dimension $m+1$, $m > b \geq 0$, having the property

$$(0^j \quad 1^j \quad \cdots \quad m^j)\boldsymbol{\lambda} = 0, \qquad j = 0, 1, \ldots, b. \tag{1.3.4}$$

Then we call $\boldsymbol{\lambda}$ a generalised increment-vector (GIV) of order b, which we denote by $\boldsymbol{\lambda} \in \Lambda_b$. For example, $\boldsymbol{\lambda} = (1 \quad -2 \quad 1)' \in \Lambda_1$; since $(1 \quad 1 \quad 1)\boldsymbol{\lambda} = 0$ and $(0 \quad 1 \quad 2)\boldsymbol{\lambda} = 0$. Clearly (1.3.4) is the vector analogue of (1.3.1) when $n = 1$.

From (1.2.9), we see immediately that, for all $b \in \mathcal{N}$,

$$\boldsymbol{\pi}_b \stackrel{\text{def}}{=} \left(1 \quad -\binom{b+1}{1} \quad \cdots \quad (-1)^i\binom{b+1}{i} \quad \cdots \quad (-1)^{b+1}\right)' \in \Lambda_b; \tag{1.3.5}$$

and we call such a $\boldsymbol{\pi}_b$ the fundamental increment vector (FIV) of order b. According to (1.2.10), on choosing $m = b+1$, we can write

$$\boldsymbol{\pi}_b = A\{(1-B)^{b+1}\} = A\{\nabla^{b+1}\}. \tag{1.3.6}$$

If $\boldsymbol{\lambda} \in \Lambda_b$, then, for any real t and an integer $j \in [0, b]$,

$$((t \pm 0)^j \quad \cdots \quad (t \pm m)^j)\boldsymbol{\lambda} = \sum_{i=0}^{j}(\pm 1)^i \binom{j}{i} t^{j-i}(0^i \quad \cdots \quad m^i)\boldsymbol{\lambda} = 0, \tag{1.3.7a}$$

using (1.3.4), and everywhere making the same choice from the $\pm$. Hence, if $f(t)$ is any polynomial in t of degree not greater than b, we have (again choosing just one sign throughout)

$$(f(t \pm 0) \quad \cdots \quad f(t \pm m))\boldsymbol{\lambda} = 0 \tag{1.3.7b}$$

for all real t; and we say that the vector $\boldsymbol{\lambda}$ annihilates the polynomial $f(t)$.

The definition of the GIV here is substantially the same as it was in Cressie (1988, relation 3.4); but it has the advantage (for our purposes) of avoiding the complications that arise from the involvement of the series length, and the possibility of missing observations (which Cressie considered).

Note that there are no GIV of order b with dimension $b+1$ or less. For convenience (see Theorems 1.3.2 and 1.3.3, below), we denote the set of all vectors (of any finite dimension) by Λ_{-1}. Then, from (1.3.4), it is evident that

$$\Lambda_{-1} \supset \Lambda_0 \supset \Lambda_1 \supset \cdots. \tag{1.3.8}$$

Thus, we may write

$$\tilde{\Lambda}_b = \Lambda_b \cap \bar{\Lambda}_{b+1}, \qquad b \in \mathcal{N} \cup \{-1\}, \tag{1.3.9}$$

where $\bar{\Lambda}_{b+1}$ is the complement of Λ_{b+1}; and then we have

$$\boldsymbol{\pi}_b \in \tilde{\Lambda}_b, \tag{1.3.10}$$

due to (1.3.5) and the fact that (1.2.9) gives $(0^{b+1} \quad 1^{b+1} \quad \cdots \quad (b+1)^{b+1})\boldsymbol{\pi}_b \neq 0$.

1.3.4 The GIV decomposition theorem

We first introduce the notation

$$\boldsymbol{\pi}_b^{(i,\,j)} \overset{\text{def}}{=} (\overbrace{0 \quad \cdots \quad 0}^{i\ zeros} \quad \boldsymbol{\pi}_b' \quad \overbrace{0 \quad \cdots \quad 0}^{j\ zeros})'. \tag{1.3.11}$$

Theorem 1.3.1 *Consider any $\boldsymbol{\lambda}$ of finite dimension $m+1$, $m \geq b+1$. Then $\boldsymbol{\lambda} \in \Lambda_b$ if and only if $\boldsymbol{\lambda}$ can be decomposed as*

$$\boldsymbol{\lambda} = \sum_{i=0}^{m-b-1} \alpha_i \boldsymbol{\pi}_b^{(i,\, m-b-1-i)}. \tag{1.3.12}$$

Proof "Only if". First consider $m = b+1$. In view of (1.3.4), any arbitrary GIV of dimension $b+2$ and order b, $\boldsymbol{\lambda} = (\lambda_0 \cdots \lambda_{b+1})'$ say, must satisfy

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & b+1 \\ \vdots & \vdots & & \vdots \\ 0^b & 1^b & \cdots & (b+1)^b \end{pmatrix} \begin{pmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_{b+1} \end{pmatrix} = \mathbf{0}, \tag{1.3.13}$$

which can be written equivalently as

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & b+1 \\ \vdots & \vdots & & \vdots \\ 1 & 2^b & \cdots & (b+1)^b \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{b+1} \end{pmatrix} = \begin{pmatrix} -\lambda_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{1.3.14}$$

The Vandermonde matrix, on the left of (1.3.14) is necessarily non-singular (cf Lakshmikantham and Trigiante 1988, p 35); and so (1.3.14) has a unique solution for $(\lambda_1 \quad \cdots \quad \lambda_{b+1})$ when λ_0 is given. Thus, if we replace λ_0 by $\tilde{\lambda}_0 = \alpha\lambda_0$, the solution modifies to $(\tilde{\lambda}_1 \quad \cdots \quad \tilde{\lambda}_{b+1}) = (\alpha\lambda_1 \quad \cdots \quad \alpha\lambda_{b+1})$. In particular, the FIV of order b, $\boldsymbol{\pi}_b$, has all of its elements proportional to those of the originally chosen arbitrary GIV — and (1.3.12) holds.

Now suppose $m > b+1$. Then it is easy to see that we may write

$$\boldsymbol{\lambda}' = (\tilde{\lambda}_0 \cdots \tilde{\lambda}_{b+1}\ 0 \cdots 0) + \alpha_1(0 \quad \boldsymbol{\pi}_b' \quad 0 \cdots 0) + \cdots + \alpha_{m-b-1}(0 \cdots 0 \quad \boldsymbol{\pi}_b'), \quad (1.3.15)$$

where $\tilde{\lambda}_0, \ldots, \tilde{\lambda}_{b+1}$ and $\alpha_1, \ \ldots \ , \alpha_{m-b-1}$ are appropriate constants. Thus, in view of (1.3.4),

$$\begin{pmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ 0 & 1 & \cdots & b+1 & \cdots & m \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 1 & \cdots & (b+1)^b & \cdots & m^b \end{pmatrix} \left\{ \begin{pmatrix} \tilde{\lambda}_0 \\ \vdots \\ \tilde{\lambda}_{b+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \alpha_1 \begin{pmatrix} 0 \\ \boldsymbol{\pi}_b \\ 0 \\ \vdots \\ \\ 0 \end{pmatrix} + \cdots + \alpha_{m-b-1} \begin{pmatrix} 0 \\ \\ \vdots \\ \\ 0 \\ \boldsymbol{\pi}_b \end{pmatrix} \right\}$$
$$= \mathbf{0}. \quad (1.3.16)$$

Next, because of (1.3.7a), any null vector of dimension $m+1$, with $b+2$ of its consecutive elements replaced by $\boldsymbol{\pi}_b$, is orthogonal to all the rows of the premultiplying $(b+1) \times (m+1)$ matrix on the left of (1.3.16). So

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & b+1 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & (b+1)^b \end{pmatrix} \begin{pmatrix} \tilde{\lambda}_0 \\ \tilde{\lambda}_1 \\ \vdots \\ \tilde{\lambda}_{b+1} \end{pmatrix} = \mathbf{0},$$

which is precisely of the form (1.3.13). Consequently,

$$(\tilde{\lambda}_0 \ \cdots \ \tilde{\lambda}_{b+1}) = \alpha_0 \boldsymbol{\pi}_b';$$

and, substituting this into (1.3.15), then shows that (1.3.12) again holds.

"If". Since $\boldsymbol{\pi}_b \in \Lambda_b$, for $i = 0, 1, \ldots, m-b-1$, (1.3.7a) gives (on putting $t = i$ and $m = b+1$ there):

$$(i^j \ \cdots \ (i+b+1)^j)\boldsymbol{\pi}_b = 0$$
$$\Rightarrow (0^j \ \cdots \ i^j \ \cdots \ (i+b+1)^j \cdots m^j)\boldsymbol{\pi}_b^{(i,\ m-b-1-i)} = 0;$$

and the truth of (1.3.12) implies that $(0^j \ \cdots \ m^j)\boldsymbol{\lambda} = 0$ which, due to (1.3.4), ensures that $\boldsymbol{\lambda} \in \Lambda_b$. □

The following corollaries of Theorem 1.3.1 will be found useful. (The dot denotes that the order of the vector's elements has been reversed.)

Corollary 1.3.1 *The vector* $\boldsymbol{\lambda}$, $(\lambda_0 \ \ldots \ \lambda_m)'$ *say,* $\in \Lambda_b$ *if and only if*

$$\dot{\boldsymbol{\lambda}} = (\lambda_m \ \cdots \ \lambda_0)' \in \Lambda_b.$$

Proof By Theorem 1.3.1, $\boldsymbol{\lambda} \in \Lambda_b$ if and only if (1.3.12) holds; i.e., if and only if

$$\dot{\boldsymbol{\lambda}} = \sum_{i=0}^{m-b-1} \alpha_i \dot{\boldsymbol{\pi}}_b^{(i,\ m-b-1-i)} = (-1)^{b+1} \sum_{i=0}^{m-b-1} \alpha_i \boldsymbol{\pi}_b^{(m-b-1-i,\ i)};$$

that is, again by Theorem 1.3.1, if and only if $\dot{\boldsymbol{\lambda}} \in \Lambda_b$. □

Corollary 1.3.2 *If* $\boldsymbol{\lambda} \in \Lambda_b$ *has dimension* n [illegible] 1, *and* $f(t)$ *is a proper polynomial of degree* a, *then (either using* $+$ *throughout* [illegible] *throughout)*

$$f_{\pm}^{\boldsymbol{\lambda}}(t) \stackrel{\text{def}}{=} (f(t \pm 0) \ \cdots \ f(t \pm m))\boldsymbol{\lambda} \tag{1.3.17}$$

are proper polynomials of degree $a - b - 1$ *(where, when* $a \leq b$, *this means that the* $f_{\pm}^{\boldsymbol{\lambda}}(t)$ *are null, in agreement with (1.3.7b)).*

Proof From (1.3.12),

$$\begin{aligned} f_{\pm}^{\boldsymbol{\lambda}}(t) &= \sum_{i=0}^{m-b-1} \alpha_i (f(t \pm 0) \cdots \ f(t \pm m))\boldsymbol{\pi}_b^{(i,\ m-b-1-i)} \\ &= \sum_{i=0}^{m-b-1} \alpha_i (f(t \pm i) \ \cdots \ f(t \pm i \pm b \pm 1))\boldsymbol{\pi}_b \end{aligned}$$

which is, on using (1.2.8),

$$\sum_{i=0}^{m-b-1} [\alpha_i \times \text{a proper polynomial (of degree } a - b - 1, \text{if } a > b; \text{ or null, if } a \leq b)]. \quad \Box$$

1.3.5 I_d and $\tilde{I}_d$ series

The rough idea of using d in model (1.1.1) to classify integrated stationary time series is well-known. However, to make increment-vector methodology work, we need more rigorous definitions.

From the point of view of sampling processes in continuous time, Cressie (1988) coined the term *I_d-process*. We prefer to call this an *I_d-series* and provide a fresh definition.

Let the set, I_0, be comprised of all second-order stationary series defined on a certain probability space $(\Omega, \mathcal{F}, P)$. Then, for any $d \in \mathcal{Z}^+$, we call $\{Z(t) : t = 0, 1, \ldots\}$ an *I_d-series* if and only if

$$\{\nabla^d Z(t)\} \in I_0, \tag{1.3.18}$$

and denote the set of all I_d-series by I_d. Also, if $\{\nabla^{d-1} Z(t)\} \in I_0$, then evidently $\{\nabla^d Z(t)\} \in I_0$. Thus, we have

$$I_0 \subset I_1 \subset I_2 \subset \cdots. \tag{1.3.19}$$

Next, using the notation of (1.3.17), write

$$Z_-^{\boldsymbol{\lambda}}(t) = (Z(t) \ \cdots \ Z(t-m))\boldsymbol{\lambda} = A^{-1}(\boldsymbol{\lambda})Z(t), \quad t = m, m+1, \ldots . \tag{1.3.20}$$

Then the following theorem connects the concepts of the GIV and I_d-series.

Theorem 1.3.2 $\{Z(t)\} \in I_d, d \in \mathcal{N}$, *if and only if for all* $\boldsymbol{\lambda} \in \Lambda_{d-1}$

$$\{Z_-^{\boldsymbol{\lambda}}(t)\} \in I_0. \tag{1.3.21}$$

Proof Putting $b = d - 1$, (1.3.20) followed by (1.3.12) yields:

$$\begin{aligned} Z_-^{\boldsymbol{\lambda}}(t) &= A^{-1}\{\sum_{i=0}^{m-b-1} \alpha_i \boldsymbol{\pi}_b^{(i,\ m-b-1-i)}\} Z(t) = \sum_{i=0}^{m-b-1} \alpha_i A^{-1}(\boldsymbol{\pi}_b^{(i,\ m-b-1-i)}) Z(t) \\ &= (\sum_{i=0}^{m-b-1} \alpha_i B^i)\nabla^d Z(t) = \alpha(B)\nabla^d Z(t), \end{aligned} \tag{1.3.22}$$

say, which is clearly second-order stationary when and only when (1.3.18) is true. The "when" is immediate, whilst the "only when" can be achieved by choosing $\boldsymbol{\lambda} = \boldsymbol{\pi}_{d-1}$ which gives $\alpha(B) = 1$. □

Relation (1.3.21) coincides with the definition of the I_d-process given in Cressie (1988). But, evidently, (1.3.18) will be simpler to check, in general, than (1.3.21). For instance, we can immediately see that a series $\{Z(t) : t = 0, 1, \ldots\}$, satisfying (1.1.1) and (1.1.5), is an I_d-series.

Although the definition of an I_d-series is helpful, it does not uniquely classify integrated stationary series. To provide a unique classification, we can define subsets of the I_d-series, by

$$\tilde{I}_d = I_d \cap \bar{I}_{d-1}, \qquad d \in \mathcal{Z}^+, \tag{1.3.23}$$

where $\bar{I}_{d-1}$ is the complement of I_{d-1}. If $\{Z(t)\} \in \tilde{I}_d$, we call $\{Z(t)\}$ an $\tilde{I}_d$-series.

For example, consider (1.1.1) with $\{W(t)\}$ a stationary and invertible ARMA (p,q) series. Then $\{Z(t)\} \in \tilde{I}_d$. But observe that, without the restriction of invertibility, $\{Z(t)\}$ could belong to an $\tilde{I}_a$ ($a < d$; where, when $a < 1, \tilde{I}_a$ is defined below). For instance, if $W(t) = A(t) - A(t-1)$, then $\{Z(t)\} \in \tilde{I}_{d-1}$.

I_0 can be partitioned into $\tilde{I}_0$, the set of all stationary series that are not overdifferenced, and I_{-1}, another set of all the stationary series that have various positive degrees of overdifferencing [that is, $\{Z(t)\} \in I_{-1}$ if and only if there is a $\{W(t)\} \in I_0$ such that $Z(t) = \nabla W(t)$]. It is straightforward to next partition I_{-1} into $\tilde{I}_{-1}$ and I_{-2}, and so on, with the partition of a general I_{-i} into $\tilde{I}_{-i}$ and I_{-i-1}, giving $I_{-1} = \cup_{i=1}^{\infty} \tilde{I}_{-i}$, a union of disjoint sets, with $\tilde{I}_{-i}$ denoting the set of all exactly i-times overdifferenced series, and $I_{-1} \supset I_{-2} \supset \cdots$. In the following, we only use I_d ($d \geq -1$) and $\tilde{I}_d$ ($d \geq 0$).

Observe that, if the spectral density of the regular component of the Wold decomposition (see Doob 1953) of a series in I_0 is denoted by $s(\omega)$, then this series is in $\tilde{I}_0$ if and only if

$$\lim_{\omega \to 0} s(\omega)/|\, 1 - e^{-i\omega} \,|^2 \longrightarrow \infty \, . \tag{1.3.24}$$

Moreover, it follows from the definition of an I_d-series [(1.3.18)] and (1.3.23) that, when $d \in \mathcal{N}$, $\{Z(t)\} \in \tilde{I}_d$ if and only if

$$\{\nabla^d Z(t)\} \in \tilde{I}_0. \tag{1.3.25}$$

So, when $\{Z(t)\} \in \tilde{I}_d, d \in \mathcal{N}$,

$$\{\nabla^{b+1} Z(t)\} \in \begin{cases} \tilde{I}_{d-1-b}, & \text{if} \;\; b \leq d-1, \\ I_{-1}, & \text{if} \;\; b > d-1. \end{cases} \tag{1.3.26}$$

Corresponding to Theorem 1.3.2 and (1.3.26), respectively, we have the following pair of theorems, whose proofs are delayed to Section 1.6.

Theorem 1.3.3 $\{Z(t)\} \in \tilde{I}_d,\ d \in \mathcal{N}$ *if and only if for all* $\boldsymbol{\lambda} \in \tilde{\Lambda}_{d-1}$,

$$\{Z_-^{\boldsymbol{\lambda}}(t)\} \in \tilde{I}_0\,. \tag{1.3.27}$$

Theorem 1.3.4 *When* $\{Z(t)\} \in \tilde{I}_d, d \in \mathcal{N}$, *for all* $\boldsymbol{\lambda} \in \tilde{\Lambda}_b, b \in \mathcal{N}$,

$$\{Z_-^{\boldsymbol{\lambda}}(t)\} \in \begin{cases} \tilde{I}_{d-1-b}, & \text{if } b \le d-1, \\ I_{-1}, & \text{if } b > d-1. \end{cases} \tag{1.3.28}$$

It follows from Theorem 1.3.4 that, for $b \in \mathcal{N}$,

$$\{Z_-^{\boldsymbol{\lambda}}(t)\} \in \tilde{I}_0 \cup I_{-1} = I_0 \tag{1.3.29}$$

if and only if

$$b \ge \max\{d-1,\ 0\}. \tag{1.3.30}$$

When $\{Z(t)\} \in I_d$, (1.3.30) is a sufficient condition for (1.3.29).

1.4 Generalized Covariance Functions and Variograms

1.4.1 Cressie's primary increment vectors

Using the GIV, the specialization of the GCF, defined by (1.3.3), to the I_d-series case is straightforward. If $\{Z(t)\} \in I_d$, then, for any $\boldsymbol{\lambda} = (\lambda_0\ \cdots\ \lambda_m)' \in \Lambda_b$ and $\boldsymbol{\mu} = (\mu_0\ \cdots\ \mu_m)' \in \Lambda_b$ with b subject to (1.3.30), the covariance between $Z_-^{\boldsymbol{\lambda}}(t)$ and $Z_-^{\boldsymbol{\mu}}(t)$ exists and does not depend on t. In such a situation, if there exists an even function $\kappa_b(h)$, satisfying

$$\mathrm{Cov}\{\sum_{i=0}^{m} \lambda_i Z(t-i), \sum_{j=0}^{m} \mu_j Z(t-j)\} = \sum_{i=0}^{m}\sum_{j=0}^{m} \lambda_i \mu_j \kappa_b(j-i), \quad t \ge m, \tag{1.4.1}$$

we call $\kappa_b(h)$ a generalised covariance function (GCF) of order b for $\{Z(t)\}$.

For the purpose of obtaining $\kappa_b(h)$, Cressie (1988) introduced the following special GIVs, of order b and dimension $h+b+1$, which he termed primary increment-vectors (PIV),

$$\boldsymbol{\nu}_b^h = (\nu_b^h(0)\ \cdots\ \nu_b^h(h+b))', \quad h = 1,\ 2,\ \ldots, \tag{1.4.2}$$

defined for $b \in \{0,\ 1,\ 2\}$ by:

$$\boldsymbol{\nu}_0^h = (1 \quad 0 \ \cdots \ 0 \ -1)' \tag{1.4.3a}$$

$$\boldsymbol{\nu}_1^h = (1 \quad 0 \ \cdots \ 0 \ -(h+1) \quad h)' \tag{1.4.3b}$$

$$\boldsymbol{\nu}_2^h = (2 \quad 0 \ \cdots \ 0 \quad -(h+1)(h+2) \quad 2h(h+2) \quad -h(h+1))'. \tag{1.4.3c}$$

Later on, we will extend the definition of these early PIVs to any $b \in \mathcal{N}$ and will find that it is more convenient to define a PIV as

$$\boldsymbol{\pi}_b^h = \boldsymbol{\nu}_b^h / b!\ , \qquad h = 1,\ 2,\ \ldots, \tag{1.4.4}$$

which ensures the first element is always unity. For $b = 0$ and 1, $\boldsymbol{\pi}_b^h = \boldsymbol{\nu}_b^h$. Comparing (1.4.4) with (1.3.5),we see that

$$\boldsymbol{\pi}_b^1 = \boldsymbol{\pi}_b, \qquad b \in \{0,\ 1,\ 2\}. \tag{1.4.5}$$

1.4.2 Variograms of general order

The variogram (of order 0), defined by (1.1.9) and (1.1.10), is well-known in the literature. Its generalization will offer an alternative to the GCF as a moment description for an I_d-series.

For a stationary series, $\{W(t)\}$, we may define a family of stationary series $\{X(h,t) = W(t)W(t+h) : t = \ldots, -1, 0, 1, \ldots\}, h = 0, 1, \ldots$. Then the ACVF of $\{Z(t)\}, \gamma(h)$, is defined by $E\{X(h,t)\}$ which does not depend on t. For an $\tilde{I}_d$-series, $\{Z(t)\}$, and any integer b satisfying (1.3.30), instead of producing just a single stationary series, ${}_1\nabla^{b+1}Z(t)\}$ (corresponding to $\boldsymbol{\pi}_b^1$), we use the whole sequence of PIVs from $\tilde{\Lambda}_b$, $\{\boldsymbol{\pi}_b^h : \ h = 1, 2, \ldots\}$, and produce the family of stationary series, $[\{Y_b(h+b,t)\} : \ h = 1, 2, \ldots](t > b)$, given by:

$$Y_b(h+b,t) = \mathbf{Z}'_{h+b}(t)\boldsymbol{\pi}_b^h = \sum_{i=0}^{h+b} \pi_b^h(i) Z(h+b-i+t), \tag{1.4.6}$$

where

$$\mathbf{Z}_{h+b}(t) = (Z(h+b+t) \ Z(h+b-1+t) \ \cdots \ Z(t))', \quad t = 0, 1, \ldots . \tag{1.4.7}$$

Then, any moment of $Y_b(h+b,t)$ which does not depend on h, say

$$v_b(h) = \mathrm{Var}\{Y_b(h+b,t)\} = E\{Y_b^2(h+b,t)\}, \quad h = 1,2,\ldots, \tag{1.4.8}$$

may be used to characterize $\{Z(t)\}$. In fact, one could also consider alternative sequences of vectors from Λ_b rather than $\{\boldsymbol{\pi}_b^h : h = 1,2,\ldots\}$ or, instead of a variance as in (1.4.8), use cross-covariances between family-member series [as was done by Yaglom 1962, p 88; 1987, p 427]. However, the choice of (1.4.6) and (1.4.8) has merits which will be seen later in this thesis. Also, the variogram defined by (1.1.9) and (1.1.10) is a special case ($b = 0$) of (1.4.6) and (1.4.8). In consequence, we call $v_b(h)$ the *variogram of order b*.

1.4.3 Equations connecting the GCF and variogram of order b

In the spirit of Cressie (1988), we state a general approach for deriving a difference equation that the GCF must satisfy.

Taking $\boldsymbol{\lambda} = \boldsymbol{\mu} = \boldsymbol{\pi}_b^h$, (1.4.8) and (1.4.6) and then (1.4.1) give

$$v_b(h) = \mathrm{Var}\{\sum_{i=0}^{h+b} \pi_b^h(i) Z(h+b-i+t)\} = (\boldsymbol{\pi}_b^h)' K_{h+b} \boldsymbol{\pi}_b^h, \quad h = 1,2,\ldots, \tag{1.4.9}$$

with

$$K_{h+b} = \begin{pmatrix} \kappa_b(0) & \cdots & \kappa_b(h+b) \\ \vdots & \ddots & \vdots \\ \kappa_b(h+b) & \cdots & \kappa_b(0) \end{pmatrix}, \tag{1.4.10}$$

where (1.4.9) provides an equation for the unknown function, $\kappa_b(h)$.

Due to the string of $h-1$ consecutive elements, $\pi_b^h(1), \ldots, \pi_b^h(h-1)$, being zero and $\pi_b^h(0) = 1$ when $b = 0, 1$, or 2 [see (1.4.3) and (1.4.4)], (1.4.8) then reduces to

$$v_b(h) = (1 \;\; \pi_b^h(h) \;\; \cdots \;\; \pi_b^h(h+b)) \begin{pmatrix} \kappa_b(0) & \kappa_b(h) & \cdots & \kappa_b(h+b) \\ \kappa_b(h) & \kappa_b(0) & \cdots & \kappa_b(b) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_b(h+b) & \kappa_b(b) & \cdots & \kappa_b(0) \end{pmatrix} \begin{pmatrix} 1 \\ \pi_b^h(h) \\ \vdots \\ \pi_b^h(h+b) \end{pmatrix},$$

$$h = 1,2,\ldots. \tag{1.4.11}$$

So

$$v_b(h) = \kappa_b(0) + 2\{\pi_b^h(h)\kappa_b(h) + \cdots + \pi_b^h(h+b)\kappa_b(h+b)\}$$
$$+(\pi_b^h(h)\cdots\pi_b^h(h+b))\begin{pmatrix}\kappa_b(0) & \cdots & \kappa_b(b)\\ \vdots & \ddots & \vdots\\ \kappa_b(b) & \cdots & \kappa_b(0)\end{pmatrix}\begin{pmatrix}\pi_b^h(h)\\ \vdots\\ \pi_b^h(h+b)\end{pmatrix}, \quad h = 1, 2, \ldots, \tag{1.4.12a}$$

which has the form

$$v_b(h) = 2\{\pi_b^h(h)\kappa_b(h) + \cdots + \pi_b^h(h+b)\kappa_b(h+b)\}$$
$$+ \text{ terms in } h \text{ and } \kappa_b(0), \ldots, \kappa_b(b). \tag{1.4.12b}$$

Divide both sides of (1.4.11) by $(-1)^{b+1}2(h+b)^{(b+1)}$. Then, writing

$$\beta_b(h) = b!(-1)^{b+1}v_b(h)/\{2(h+b)^{(b+1)}\}, \quad h = 1, 2, \ldots, \tag{1.4.13}$$

and

$$\xi_b(h) = \kappa_b(h)/h, \quad h = 1, 2, \ldots, \tag{1.4.14}$$

and using (1.4.3) and (1.4.4), (1.4.12b) becomes

$$\beta_b(h) = \Delta^b\xi_b(h) + \text{ terms in } h \text{ and } \kappa_b(0), \ldots, \kappa_b(b), \quad h = 1, 2, \ldots, \tag{1.4.15}$$

for $b = 0, 1,$ and 2 (subject to $b \geq d - 1$). We may then obtain a solution for $\xi_b(h)$ in terms of $b + 1$ arbitrary constants, $\kappa_b(0), \ldots, \kappa_b(b)$, and hence find $\kappa_b(h)$ as well. [Of course, for $b = 0$, this is overkill, as we can immediately see from (1.4.12a) that $v_0(h) = 2\kappa_0(0) - 2\kappa_0(h)$.]

Notice that the difference equation of order b, (1.4.15), starts from $h = 1$, so it is natural to refer to $\kappa_b(1), \ldots, \kappa_b(b)$ as the *initial values.* We may also call $\kappa_b(0)$ an initial value for convenience.

If we specialise, by putting all these $b + 1$ initial values to zero, and denote the resulting $\xi_b(h)$ by $\tilde{\xi}_b(h)$ [and the corresponding $\kappa_b(h)$ by $\tilde{\kappa}_b(h)$], (1.4.15) then reduces to

$$\beta_b(h) = \Delta^b\tilde{\xi}_b(h), \quad h = 1, 2, \ldots, \tag{1.4.16}$$

for b satisfying $\max\{d-1, 0\} \leq b \leq 2$. It is evident, using Lemma 1.2.2 (with $r = 1$), that a particular solution of this reduced difference equation is

$$\tilde{\kappa}_b(h) = h\tilde{\xi}_b(h) = h\Sigma^b\beta_b(h), \quad h = 1, 2, \ldots . \tag{1.4.17}$$

That is, trivially for $b = 0$,

$$\tilde{\kappa}_0(h) = h\tilde{\xi}_0(h) = h\beta_0(h) = -v_0(h)/2, \quad h = 1, 2, \ldots \tag{1.4.18}$$

with $\tilde{\kappa}_0(0) = 0$; but, more interestingly for $b = 1$ and 2,

$$\tilde{\kappa}_b(h) = h\tilde{\xi}_b(h) = \frac{h}{(b-1)!}\sum_{i=1}^{h-b}(h-1-i)^{(b-1)}\beta_b(i), \quad h = b+1, b+2, \ldots, \tag{1.4.19}$$

with $\tilde{\kappa}_b(0)$ and $\tilde{\kappa}_b(1), \ldots, \tilde{\kappa}_b(b)$ all zero.

In fact, all the formulae in this subsection are true for any b and d in $\mathcal{N}$ satisfying (1.3.30), if we adopt the definition for the PIV of general order b, as is given in the next section (see Corollary 1.5.2).

Notice that, at present, we only call the $\tilde{\kappa}_b(h)$ in (1.4.19) a *particular* solution of (1.4.9). We have not claimed that it is a GCF, since we have not yet verified that $\tilde{\kappa}_b(h)$ satisfies (1.4.1) for all $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ in Λ_b. We will discuss this point in Section 2.1.

1.4.4 Polyvariograms

For $b = 0, 1$ and 2, we may now define two types of polyvariogram of order b as

$$\gamma_b(h) = v_b(h)/h^{2b}, \quad h = 1, 2, \ldots, \tag{1.4.20}$$

and

$$\gamma_b^*(h) = (-1)^{b+1}\tilde{\kappa}_b(h)/h^{2b}, \quad h = b+1, b+2, \ldots. \tag{1.4.21}$$

Evidently

$$\gamma_0(h) = v_0(h), \quad \gamma_0^*(h) = v_0(0)/2, \quad h = 1, 2, \ldots. \tag{1.4.22}$$

Note that $\gamma_0^*(h)$ is Cressie's (1988) semivariogram, and it can be shown that $\gamma_1^*(h)$ is his linvariogram while $6\gamma_2^*(h)$ is his quadvariogram. So, we may refer to $\gamma_b^*(h)$ as the *modified Cressie polyvariogram*. For higher $b, b \in \mathcal{N}$, (1.4.20) and (1.4.21) are still available, if we define the PIV of general order b as in the next section.

1.5 The Primary Increment Vectors of General Order b

1.5.1 The form of a general-order PIV and Cressie's proposal

We have seen that, for $b \in \{0,1,2\}$, each PIV given by (1.4.3) plays an important role in the derivation of the corresponding GCF, $\kappa_b(h)$, for an $\tilde{I}_d$–series, when $b \geq \max\{d-1,\ 0\}$, as the PIV can then be used to obtain a difference equation that the $\kappa_b(h)/h$ must satisfy. With the prospect of obtaining higher-order GCF, we now determine the PIV of general-order $b \in \mathcal{N}$. This general-order PIV will also find other fundamental applications in the representation and decomposition of $\tilde{I}_d$-series (Chapter 3) and in polyvariogram methodology (Chapter 4). Scrutinising the low order definitions, (1.4.3), and the way the corresponding difference equations for $\kappa_b(h)/h$ were derived [subsection 1.4.3 down to (1.4.15)], we see that the crucial features of the low order $\boldsymbol{\pi}_b^h$ are:

(i) The string of $h-1$ zeros, following the leading element; and

(ii) The indication of structure in the remaining $b+1$ non-zero elements.

Cressie (1988), aware of these patterns, suggested that one consider the $(h+b)$-th degree polynomial operator,

$$b!(1-B)^{b+1}\{\sum_{i=1}^{h}(\sum_{j=1}^{i} j^{b-1})B^{i-1}\}, \quad b \in \mathcal{Z}^+; \tag{1.5.1}$$

the k-th element of $\boldsymbol{\nu}_b^h$, $\nu_b^h(k)$ $(k = 0,\ldots,h+b)$, being then given by the coefficient of B^k in (1.5.1). This statement can be condensed using the assign notation, viz:

$$\boldsymbol{\nu}_b^h = b!A\{(1-B)^{b+1}\sum_{i=1}^{h}(\sum_{j=1}^{i} j^{b-1})B^{i-1}\}, \quad b \in \mathcal{Z}^+. \tag{1.5.2}$$

It is easy to check that, for $b = 1$ or 2, (1.5.2) retrieves (1.4.3). But, for higher b, (1.5.2) gives disagreement with the "$h-1$ zeros" property. For instance, when $b = 3$, (1.5.2) gives $\nu_3^h(1) = 6$ (rather that zero); while, for $b = 4$, we get the non-zero $\nu_4^h(1) = 96$ and $\nu_4^h(2) = 24$.

When we have merely the additional $\pi_b^h(1) = \nu_b^h(1)/b! \neq 0, \kappa_b(h-1)$ must also appear in the relevant difference equation [corresponding to (1.4.12a)], raising its order by one; and *a priori* it seems that the general difference equation, resulting from (1.5.2), could be much more complicated to solve than (1.4.12a), given $b \geq 3$. We will therefore consider a different definition for the general order PIV, which appears a more natural choice, and allows (1.4.12a) to be retained for all $b \geq 0$.

1.5.2 A preferred general-order PIV

For any $b \in \mathcal{N}$, write

$$1 = (1-B)^{b+1}(1-B)^{-b-1} = (1-B)^{b+1}Q(B) + R(B), \tag{1.5.3}$$

where the "quotient" $Q(B)$ is the initial part of the expansion of $(1-B)^{-b-1}$, up to and including the term in B^{h-1}, and the "remainder" $R(B)$ is a polynomial in B whose coefficients of B^0 through B^{h-1} are all zero. Now $(1-B)^{b+1}Q(B)$ is a polynomial of degree $m = h + b$, so (1.5.3) gives:

$$(1-B)^{b+1}Q(B) = 1 - R(B) = 1 - \sum_{i=0}^{b} a_i B^{h+i}, \tag{1.5.4}$$

for some $b+1$ constants, $a_0, \ldots, a_b$. Then, for the PIV of general order b, we propose

$$\boldsymbol{\pi}_b^h \stackrel{\text{def}}{=} A\{(1-B)^{b+1}Q(B)\} \tag{1.5.5}$$

which is a vector of length $h + b + 1$ with the required string of zero elements.

Now, for any $b \in \mathcal{N}$, the negative binomial expansion gives:

$$(1-B)^{-b-1} = \sum_{i=0}^{\infty} \frac{(b+1)\ \cdots\ (b+i)}{i!} B^i = \frac{1}{b!}\sum_{i=0}^{\infty}(b+i)^{(b)}B^i. \tag{1.5.6}$$

So, clearly,

$$Q(B) = \frac{1}{b!}\sum_{i=0}^{h-1}(b+i)^{(b)}B^i. \tag{1.5.7}$$

Then, substituting (1.5.7) into (1.5.5), we get

$$b!\boldsymbol{\pi}_b^h = A\{(1-B)^{b+1}\sum_{i=1}^{h}(b-1+i)^{(b)}B^{i-1}\}, \quad b \in \mathcal{N}, \tag{1.5.8}$$

where we have rewritten the summation in an alternative form which provides a clearer comparison with (1.5.2).

For $b = 1$ or 2, the *actual* equivalence between the right hand sides of (1.5.2) and (1.5.8) is straightforwardly established, and both reduce to (1.4.3). Also (1.5.8) again reduces to (1.4.3) when $b = 0$, but (1.5.2) does not. [Indeed, the reductions of our $b!\pi_b^h$ to (1.4.3) become self-evident on obtaining the alternative representation for π_b^h of (1.5.11), below.] Our suggested PIV thus has precisely the right form, but we still need to demostrate that it is indeed a GIV. In fact, we show in Section 1.6 (Theorem 1.6.2) that

$$\pi_b^h \in \tilde{\Lambda}_b, \quad h = 1, 2, \dots , \tag{1.5.9}$$

which generalizes (1.3.10).

1.5.3 An alternative representation for the proposed PIV

The special case of (1.5.8), when $h = 1$, clearly gives the first PIV of order b and dimension $b+2$ as [cf (1.4.5)]

$$\pi_b^1 = A(\nabla^{b+1}) = \pi_b, \quad b \in \mathcal{N}. \tag{1.5.10}$$

However, the form of (1.4.3) perhaps suggests a generalisation:

$$\pi_b^h = A\{1 - \frac{1}{b!}\sum_{i=0}^{b}(-1)^i \binom{b}{i} \frac{(h+b)^{(b+1)}}{h+i} B^{h+i}\}, \quad b \in \mathcal{N}; \tag{1.5.11}$$

which, from (1.5.5) and (1.5.3), would be valid if and only if

$$R(B) = \frac{1}{b!}\sum_{i=0}^{b}(-1)^i \binom{b}{i} \frac{(h+b)^{(b+1)}}{h+i} B^{h+i}, \quad b \in \mathcal{N}. \tag{1.5.12}$$

Theorem 1.5.1 *The remainder $R(B)$ in (1.5.3) is also given by (1.5.12).*

Proof From (1.5.3),

$$(1-B)^{-b-1}R(B) = (1-B)^{-b-1} - Q(B) = \frac{1}{b!}B^h \sum_{i=0}^{\infty}(h+b+i)^{(b)}B^i, \tag{1.5.13}$$

using (1.5.6) and (1.5.7). Then (1.5.4), with (1.5.13), gives

$$B^{-h}R(B) = \sum_{i=0}^{b} a_i B^i = \frac{1}{b!}(1-B)^{b+1}\sum_{i=0}^{\infty}(h+b+i)^{(b)}B^i.$$

Expanding $(1-B)^{b+1}$ and comparing coefficients of B^i, we get

$$a_i = \frac{1}{b!}\sum_{j=0}^{i}(-1)^j \binom{b+1}{j} (h+b+i-j)^{(b)}, \quad b \in \mathcal{N},\ i=0,\ldots,b.$$

So, to verify the correctness of (1.5.12), we must demonstrate that

$$\sum_{j=0}^{i}(-1)^j \binom{b+1}{j} (h+b+i-j)^{(b)} = (-1)^i \binom{b}{i} \frac{(h+b)^{(b+1)}}{h+i},$$

$$b \in \mathcal{N},\ i=0,\ldots,b. \tag{1.5.14}$$

For any $b \geq 0$, (1.5.14) is clearly true for $i=0$, when both sides reduce to $(h+b)^{(b)}$. Again, for any $b \geq 0$, assume the result holds for some $i \geq 0$ $(i \leq b-1)$. Then, for $i+1$ $(\leq b)$:

$$\begin{aligned}
&\sum_{j=0}^{i+1}(-1)^j \binom{b+1}{j} \{h+b+(i+1)-j\}^{(b)} \\
&= \sum_{j=0}^{i}(-1)^j \binom{b+1}{j} \{(h+1)+b+i-j\}^{(b)} + (-1)^{i+1}\binom{b+1}{i+1}(h+b)^{(b)} \\
&= (-1)^i \binom{b}{i} \frac{(h+b+1)^{(b+1)}}{h+1+i} + (-1)^{i+1}\binom{b+1}{i+1}\frac{(h+b)^{(b+1)}}{h} \\
&= (-1)^i \frac{(h+b)^{(b+1)}}{h}\left\{\frac{b!}{i!(b-i)!}\,\frac{h+b+1}{h+i+1} - \frac{(b+1)!}{(i+1)!(b-i)!}\right\} \\
&= (-1)^i \frac{(h+b)^{(b+1)}}{h(b-i)}\binom{b}{i+1}\left\{\frac{(i+1)(h+b+1)}{h+i+1} - (b+1)\right\} \\
&= (-1)^i \frac{(h+b)^{(b+1)}}{h(b-i)}\binom{b}{i+1}\frac{h(i-b)}{h+i+1} = (-1)^{i+1}\binom{b}{i+1}\frac{(h+b)^{(b+1)}}{h+i+1}. \qquad \square
\end{aligned}$$

Note that we do not really need to restrict i to $i \leq b$; as, for $i > b$, both sides of (1.5.14) yield zero.

Corollary 1.5.1 *If we denote the PIVs of order b, $b \in \mathcal{N}$, by $\boldsymbol{\pi}_b^h = (\pi_b^h(0) \cdots \pi_b^h(h+b))$, $h = 1,2,\ldots,$ then it follows from (1.5.11) that*

$$\pi_b^h(0) = 1, \tag{1.5.15a}$$

$$\pi_b^h(k) = 0, \quad k \in \{1, \ldots, h-1\}, \tag{1.5.15b}$$

$$\pi_b^h(h+i) = (-1)^{i+1} \binom{b}{i} (h+b)^{(b+1)}/\{b!(h+i)\}, \quad i \in \{0, \ldots, b\}. \tag{1.5.15c}$$

In Section 1.4, we derived equation (1.4.15) satisfied by $\xi_b(h) = \kappa_b(h)/h$ ($b \in \{0,1,2\}$). Now, for general $b \in \mathcal{N}$, we have:

Corollary 1.5.2 *If $\{Z(t)\} \in \tilde{I}_d$ has $\kappa_b(h)$ for a GCF of order b ($b \geq \max\{d-1, 0\}$), then $\xi_b(h) = \kappa_b(h)/h$ satisfies (1.4.15).*

Proof Using (1.5.15c) to replace the $\pi_b^h(h+i)$ and then (1.2.2), we have

$$h\pi_b^h(h)\xi_b(h) + \cdots + (h+b)\pi_b^h(h+b)\xi_b(h+b)$$

$$= \frac{(h+b)^{(b+1)}}{b!} \sum_{i=0}^{b} (-1)^{i+1} \binom{b}{i} \xi_b(h+i) = \frac{(h+b)^{(b+1)}}{b!} (-1)^{b+1} \Delta^b \xi_b(h).$$

Then, (1.4.12b) and (1.4.13) give (1.4.15) for all b satisfying (1.3.30). □

1.6 Generalized Differencing Operators

Corollary 1.3.2 showed that, if a vector λ was a GIV of order b, then λ annihilated all polynomials of degree b or less. This is really the property that motivated the choice of subscript b in Λ, rather than $b+1$ [which would seem more natural if we focused on say (1.3.5), the definition of the FIV, which also reflects the relation between the FIV of order b and the $(b+1)$-times differencing operator — see (1.3.6)].

The choice of b instead of $c = b+1$ can be a source of confusion and, to minimise this, perhaps a more appropriately descriptive term than GIV of order b would be *general annihilating vector* of order b (which annihilates all polynomials of degree up to and including b). If $c = b+1$ were chosen instead of b, then we would propose the term *generalized differencing vector* of order c and symbolically denote the set of such vectors by Λ_c^* (i.e., $\Lambda_c^* = \Lambda_b$).

Should we prefer to rewrite the theory in terms of operators rather than vectors, then we would propose the term *generalized differencing operator* (GDO) of order $c = b+1$ and get the following.

For any vector $\boldsymbol{\lambda} = (\lambda_0 \ \cdots \ \lambda_m)'$, using the assign notation,

$$A^{-1}(\boldsymbol{\lambda})Z(t) = \{\sum_{j=0}^{m} \lambda_j B^j\}Z(t). \tag{1.6.1}$$

Then, for $c = b + 1 \in \mathcal{N}$, denote the set of all GDO of order c by D_c and its subset of all those GDO which are not of order $c+1$ by $\tilde{\mathcal{D}}_c$. That is:

$$A^{-1}(\boldsymbol{\lambda}) \in \mathcal{D}_c \quad \text{iff} \quad \boldsymbol{\lambda} \in \Lambda_{c-1}, \tag{1.6.2}$$

$$A^{-1}(\boldsymbol{\lambda}) \in \tilde{\mathcal{D}}_c \quad \text{iff} \quad \boldsymbol{\lambda} \in \tilde{\Lambda}_{c-1}. \tag{1.6.3}$$

Then, in view of (1.3.8) and (1.3.9),

$$\mathcal{D}_0 \supset \mathcal{D}_1 \supset \mathcal{D}_2 \supset \cdots, \quad \tilde{\mathcal{D}}_c = \mathcal{D}_c \cap \bar{\mathcal{D}}_{c+1}, \quad c \in \mathcal{N}, \tag{1.6.4}$$

where, $\bar{\mathcal{D}}_{c+1}$ denotes the complement of $\mathcal{D}_{c+1}$. Also, using (1.6.3) with (1.3.10),

$$\nabla^c = A^{-1}(\boldsymbol{\pi}_{c-1}) \in \tilde{\mathcal{D}}_c, \tag{1.6.5}$$

and ∇^c may be called the *fundamental differencing operator* (FDO) of order c.

Frequently it is more convenient to present a result in terms of the GDO rather than the GIV.

Theorem 1.6.1 *For* $c \in \mathcal{N}$, $\boldsymbol{\lambda} = (\lambda_0 \ \cdots \ \lambda_m)'$, $m \geq c+1$,

(*i*) $A^{-1}(\boldsymbol{\lambda}) \in \mathcal{D}_c$ *if and only if*

$$A^{-1}(\boldsymbol{\lambda}) = \alpha(B)\nabla^c, \tag{1.6.6}$$

where $\alpha(B)$ *is a polynomial in* B.

(*ii*) $A^{-1}(\boldsymbol{\lambda}) \in \tilde{\mathcal{D}}_c$ *if and only if (1.6.6) holds and* ∇ *does not divide* $\alpha(B)$.

Proof (*i*) From (1.6.2), (1.3.12) and (1.5.10), $A^{-1}(\boldsymbol{\lambda}) \in \mathcal{D}_c$ if and only if

$$A^{-1}(\boldsymbol{\lambda}) = \sum_{j=0}^{m-c} \alpha_j A^{-1}(\boldsymbol{\pi}_{c-1}^{(j,\, m-c-j)}) = \sum_{j=0}^{m-c} \alpha_j B^j \nabla^c. \tag{1.6.7}$$

[Cf (1.3.22).]

(*ii*) By (1.6.4), $\tilde{\mathcal{D}}_c \subset \mathcal{D}_c$ and $\tilde{\mathcal{D}}_c \not\subset \mathcal{D}_{c+1}$. So, by (*i*), $A^{-1}(\boldsymbol{\lambda}) \in \tilde{\mathcal{D}}_c$ if and only if ∇^c divides $A^{-1}(\boldsymbol{\lambda})$ but ∇^{c+1} does not. □

We can now prove the two theorems claimed in Sections 1.3.

The proof of Theorem 1.3.3 Using (1.3.22), followed by Theorem 1.6.1 (*ii*) and then (1.6.3), $Z_-^{\boldsymbol{\lambda}}(t) = \alpha(B)\nabla^d Z(t)$ with $\alpha(B)$ having no factor of ∇ if and only if $\boldsymbol{\lambda} \in \tilde{\Lambda}_{d-1}$. But, from (1.3.25), $\{Z(t)\} \in \tilde{I}_d$ if and only if $\{\nabla^d Z(t)\} \in \tilde{I}_0$; that is, if and only if the spectral density $s(\omega)$, of the regular component of $\{\nabla^d Z(t)\}$, has property (1.3.24); i.e., if and only if the spectral density $|\alpha(e^{-i\omega})|^2 s(\omega)$, of the regular component of $\{\alpha(B)\nabla^d Z(t)\}$, satisfies the corresponding property. That is, if and only if $\{Z_-^{\boldsymbol{\lambda}}(t)\} \in \tilde{I}_0$. □

The proof of Theorem 1.3.4 As for the previous proof, when $\boldsymbol{\lambda} \in \tilde{\Lambda}_b$, $Z_-^{\boldsymbol{\lambda}}(t) = \alpha(B)\nabla^{b+1}Z(t)$ with $\alpha(B)$ having no factor of ∇.

If $b > d-1$, then (1.3.26) gives $\{\nabla^{b+1}Z(t)\} \in I_{-1}$; i.e., there is a $\{W(t)\} \in I_0$, such that $\{\nabla W(t) = \nabla^{b+1}Z(t)\}$, so $Z_-^{\boldsymbol{\lambda}}(t) = \nabla\alpha(B)W(t)$. But $\{\alpha(B)W(t)\} \in I_0$, so $\{Z_-^{\boldsymbol{\lambda}}(t)\} \in I_{-1}$.

If $b \le d-1$, then $\nabla^{d-1-b}Z_-^{\boldsymbol{\lambda}}(t) = \alpha(B)\nabla^d Z(t) = Z_-^{\boldsymbol{\mu}}(t)$, where $\boldsymbol{\mu} = A\{\alpha(B)\nabla^d\} \in \tilde{\Lambda}_{d-1}$. So, since $\{Z(t)\} \in \tilde{I}_d$, Theorem 1.3.3 gives $\{Z_-^{\boldsymbol{\mu}}(t)\} \in \tilde{I}_0$ and then, from (1.3.26), $\{Z_-^{\boldsymbol{\lambda}}(t)\} \in \tilde{I}_{d-1-b}$. □

Theorem 1.6.2 *The proposed PIV of order $b, \boldsymbol{\pi}_b^h$, defined by (1.5.8) is in $\tilde{\Lambda}_b$, i.e. (1.5.9) holds for $b \in \mathcal{N}$.*

Proof In view of (1.5.5), $A^{-1}(\boldsymbol{\pi}_b^h) = \nabla^{b+1}Q(B)$, where the polynomial $Q(B)$, defined by (1.5.7), does not have ∇ as a divisor [since $Q(1) \neq 0$]. Then, by Theorem 6.1 (*ii*) and (1.6.3), $\boldsymbol{\pi}_b^h \in \tilde{\Lambda}_b$. □

Chapter 2

Second-Order Moment Characterizations

2.1 Solutions for Generalized Covariance Functions

For an $\tilde{I}_d$-series, $d \in \mathcal{N}$, when an integer b satisfies (1.3.30), then we may define its GCF of order b, $\kappa_b(h)$, by (1.4.1) — a characterization of its second-order moments. We have shown that $\kappa_b(h)$ must satisfy equation (1.4.11) or its alternative version (1.4.15) — a difference equation for $\xi_b(h) = \kappa_b(h)/h$ (see Corollary 1.5.2). The purpose of this section is to derive the general solution of (1.4.15) and to verify that the solution obtained is indeed a legitimate GCF. This GCF is a function of the variogram. As it is easy to construct estimators for the variogram, we can then get a suitable estimator for the GCF.

2.1.1 The complementary solution

Since the case $b = 0$ is trivial, we may assume $b \in \mathcal{Z}^+$. Consider the homogeneous linear difference equation with constant coefficients:

$$\Delta^b f(t) = 0, \quad t = 0, 1, \dots . \tag{2.1.1}$$

It is well-known that the solution of (2.1.1) has the form

$$f(t) = \alpha_0 + \alpha_1 t + \cdots + \alpha_{b-1} t^{b-1}, \quad t = 1, 2, \dots, \tag{2.1.2}$$

and the corresponding non-homogeneous equation (1.2.6) has the general solution

$$f(t) = \alpha_0 + \alpha_1 t + \cdots + \alpha_{b-1} t^{b-1} + \Sigma^b g(t), \quad t = 1, 2, \ldots . \tag{2.1.3}$$

Here, $\Sigma^b g(t)$ is the particular solution of (1.2.6) given by the right of (1.2.4), with $r = 1$, and the b arbitrary constants, $\alpha_0, \ldots, \alpha_{b-1}$, may be uniquely determined from b initial values of $f(t)$. (See, for instance, Hildebrand 1968, Chapter 1). We call the right of (2.1.2), in (2.1.3), the complementary solution of (1.2.6).

We need to solve the homogeneous equation corresponding to (1.4.9), viz:

$$(\boldsymbol{\pi}_b^h)' C_{h+b} \boldsymbol{\pi}_b^h = 0, \quad h = 1, 2, \ldots, \tag{2.1.4}$$

where

$$C_{h+b} = \begin{pmatrix} c(0) & c(1) & \cdots & c(h+b) \\ c(1) & c(0) & \cdots & c(h+b-1) \\ \vdots & \vdots & \ddots & \vdots \\ c(h+b) & c(h+b-1) & \cdots & c(0) \end{pmatrix}. \tag{2.1.5}$$

Consider solving the modified equation

$$(\boldsymbol{\pi}_b^h)' C^*_{h+b} \boldsymbol{\pi}_b^h = 0, \quad h = 1, 2, \ldots, \tag{2.1.6}$$

where C^*_{h+b} is just C_{h+b}, with all the elements below the main diagonal having their arguments multiplied by -1. That is, the i-th row of C^*_{h+b} can be written as

$$\mathbf{c}'(i) = (c(-i+1) \ \cdots \ c(-i+h+b+1)), \quad i = 1, \ldots, h+b+1, \tag{2.1.7}$$

and

$$C^*_{h+b} = (\mathbf{c}(1) \ \cdots \ \mathbf{c}(h+b+1))'. \tag{2.1.8}$$

Try

$$c(j) = \sum_{k=0}^{2b+1} c_k j^k, \quad j = 0, 1, \ldots, \tag{2.1.9}$$

where $c_0, \ldots, c_{2b+1}$ are arbitrary constants. That is, $c(j)$ is a polynomial in j of degree $2b+1$. Then, since $\boldsymbol{\pi}_b^h \in \Lambda_b$, if we postmultiply (2.1.7) by $\boldsymbol{\pi}_b^h$ and temporarily denote $\boldsymbol{\pi}_b^h$ by $\boldsymbol{\pi}$, we get (using the notation of Corollary 1.3.2)

$$\mathbf{c}'(i)\boldsymbol{\pi} = c_+^{\boldsymbol{\pi}}(1-i), \tag{2.1.10}$$

which is a polynomial in i of degree b (due to Corollary 1.3.2). So, from (2.1.8) and (2.1.10), we have

$$C^*_{h+b}\boldsymbol{\pi} = (c^{\boldsymbol{\pi}}_+(0) \ \cdots \ c^{\boldsymbol{\pi}}_+(-h-b))'. \tag{2.1.11}$$

Then, after transposing (2.1.11) and postmultiplying by $\boldsymbol{\pi}$, Corollary 1.3.2 gives

$$\boldsymbol{\pi}'C^*_{h+b}\boldsymbol{\pi} = 0. \tag{2.1.12}$$

Thus, we see that (2.1.9) satisfies (2.1.6).

Now, if we assume $c(j)$ is even, (2.1.6) becomes identical to (2.1.4); and then (2.1.9), the solution to (2.1.6), reduces to

$$c(j) = \sum_{i=0}^{b} a_i j^{2i}, \quad j = 0, 1, \ldots, \tag{2.1.13}$$

(where $a_i \equiv c_{2i}$) which is also a solution of (2.1.4).

Lemma 2.1.1 *Given $b+1$ initial values of $c(0), \cdots, c(b)$ [say $\tilde{c}(0), \ldots, \tilde{c}(b)$]; then, in the solution (2.1.13) of equation (2.1.4), the $b+1$ constants, $a_0, \ldots, a_b$, are uniquely determined.*

Proof From (2.1.13), the a_i clearly satisfy

$$(\tilde{c}(0) \ \tilde{c}(1) \ \cdots \ \tilde{c}(b)) = (a_0 \ a_1 \ \cdots \ a_b)V_b, \tag{2.1.14}$$

where V_b is the $(b+1) \times (b+1)$ matrix defined by

$$V_b = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1^2 & \cdots & b^2 \\ \vdots & \vdots & & \vdots \\ 0 & 1^{2b} & \cdots & b^{2b} \end{pmatrix}. \tag{2.1.15}$$

This is a Vandermonde matrix, of the form

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ v_0 & v_1 & \cdots & v_b \\ \vdots & \vdots & & \vdots \\ v_0^b & v_1^b & \cdots & v_b^b \end{pmatrix} \tag{2.1.16}$$

with $v_j = j^2$ $(j = 0, \ldots, b)$, and so is non-singular (for instance, see Lakshmikantham and Trigiante 1988, p 35). So (2.1.14) provides a unique vector of a_i's, given $b+1$ initial $c(i)$'s. □

Analogously to the rewriting of (1.4.9) as (1.4.12a), rewrite (2.1.4) as

$$0 = c(0) + 2\{\pi_b^h(h)c(h) + \cdots + \pi_b^h(h+b)c(h+b)\}$$

$$+(\pi_b^h(h) \cdots \pi_b^h(h+b))\begin{pmatrix} c(0) & \cdots & c(b) \\ \vdots & \ddots & \vdots \\ c(b) & \cdots & c(0) \end{pmatrix}\begin{pmatrix} \pi_b^h(h) \\ \vdots \\ \pi_b^h(h+b) \end{pmatrix}, \quad h = 1,2,\ldots . \quad (2.1.17)$$

Next, for $h = 1,2,\ldots$, denoting $c(h)/h$ by $e(h)$, (1.5.15c) gives (cf Corollary 1.5.2)

$$\pi_b^h(h)c(h) + \cdots + \pi_b^h(h+b)c(h+b) = \{(-1)^{b+1}/b!\}(h+b)^{(b+1)}\Delta^b e(h). \quad (2.1.18)$$

So (2.1.17) yields the following result: Given $b+1$ initial values $\tilde{c}(j)$ of c_j, $j = 0,\ldots,b$, then $e(h) = c(h)/h$ satisfies

$$\Delta^b e(h) = \frac{(-1)^b b!}{2(h+b)^{(b+1)}}\left\{\tilde{c}(0) + \begin{pmatrix} \pi_b^h(h) \\ \vdots \\ \pi_b^h(h+b) \end{pmatrix}'\begin{pmatrix} \tilde{c}(0) & \cdots & \tilde{c}(b) \\ \vdots & \ddots & \vdots \\ \tilde{c}(b) & \cdots & \tilde{c}(0) \end{pmatrix}\begin{pmatrix} \pi_b^h(h) \\ \vdots \\ \pi_b^h(h+b) \end{pmatrix}\right\}$$

$$h = 1,2,\ldots . \quad (2.1.19)$$

Theorem 2.1.1 *For an $\tilde{I}_d$-series and integers b and d satisfying (1.3.30), the complementary solution of (1.4.9) [i.e. the solution of (2.1.4)] can take the form of (2.1.13). Given the $b+1$ initial values $\tilde{c}(0),\ldots,\tilde{c}(b)$ of $c(h)$, (2.1.13) is the only solution of (2.1.4), and its $b+1$ coefficients $a_0,\ldots,a_b$, are uniquely determined by (2.1.14).*

Proof Assume that (2.1.4) has another solution, $\dot{c}(h)$ say [with, of course, the same initial values; that is $\dot{c}(h) = \tilde{c}(h)$, $h = 0,\ldots,b$]. Then, putting $\dot{e}(h) = \dot{c}(h)/h$, $h = 1,2,\ldots$, $\dot{e}(h)$ also satisfies (2.1.19). So

$$\Delta^b\{\dot{e}(h) - e(h)\} = 0, \quad h = 1,2,\ldots, \quad (2.1.20)$$

with initial values $\dot{e}(h) - e(h) = 0$, $h = 1,\ldots,b$. Now (2.1.20) is a homogeneous linear difference equation of order b, so [from (2.1.2)] its solution, $\dot{e}(h) - e(h)$, is a polynomial of degree $b-1$ at most and has zeros at b distinct values of h, and is therefore identically equal to zero. Since $\dot{c}(0) = c(0)$, it then follows immediately that

$$\dot{c}(h) = c(h), \quad h = 0,1,\ldots . \quad \square$$

Remark Strictly speaking, the uniqueness is within the class of functions that are usually referred to as the general solutions of the homogeneous linear difference equation with constant coefficients. See Lakshmikantham and Trigiante (1988, p 36). We have not considered a wider class of functions, for instance those with an additive period-1 component. See Lakshmikantham and Trigiante (1988, p 3).

2.1.2 The general solution

We have already noted the structure of the general solution of a linear difference equation [see (2.1.3)]. Let $c_b(h)$ (we now add a suffix b to emphasize the order) be the unique complementary solution of (1.4.9) [the full solution of (2.1.4)] given by (2.1.13) and (2.1.14), when its $b+1$ (arbitrary) initial values are denoted by $\tilde{c}(0),\ldots,\tilde{c}(b)$. Let $\tilde{\kappa}_b(h) = h\tilde{\xi}_b(h)$, given by (1.4.19), be a particular solution of (1.4.9) which has all its initial values, $\tilde{\kappa}_b(0),\ldots,\tilde{\kappa}_b(b)$, zero. [As mentioned in subsection 1.4.3, we also call $\tilde{c}(0)$ and $\tilde{\kappa}_b(0)$ "initial values" for convenience, although the difference equation only starts at $h=1$.] Then we have the following theorem:

Theorem 2.1.2 *For an $\tilde{I}_d$-series and integers b and d satisfying (1.3.30), the solution of (1.4.9), $\kappa_b(h)$, with its $b+1$ initial values coinciding with $b+1$ arbitrarily given constants $\tilde{c}(0),\ldots,\tilde{c}(b)$, can be uniquely expressed as*

$$\kappa_b(j) = c_b(j) + \tilde{\kappa}_b(j), \quad j = 0,1,\ldots\,. \tag{2.1.21}$$

Proof Put $\tilde{K}_{h+b}$ as (1.4.10) with the $\kappa_b(j)$ replaced by $\tilde{\kappa}_b(j)$. Then, since $\tilde{\kappa}_b(j)$ is a particular solution of (1.4.9),

$$v_b(h) = (\boldsymbol{\pi}_b^h)'\tilde{K}_{h+b}\boldsymbol{\pi}_b^h, \quad h = 1,2,\ldots\,. \tag{2.1.22}$$

So, due to (2.1.4),

$$v_b(h) = (\boldsymbol{\pi}_b^h)'(C_{h+b} + \tilde{K}_{h+b})\boldsymbol{\pi}_b^h, \quad h = 1,2,\ldots, \tag{2.1.23}$$

and (2.1.21) is the general solution of (1.4.9). This solution can be uniquely determined by putting $\kappa_b(j) = \tilde{c}(j)$, $j = 0,\ldots,b$, since the constants $a_0,\ldots,a_b$ in $c_b(j)$ [see (2.1.13)] are then uniquely given by (2.1.14). □

Example When $b=2$, follow Cressie (1988) and take $\kappa_2(0)=0$, but leave $\kappa_2(1)$ and $\kappa_2(2)$ undetermined. Then, from (2.1.13), $c_2(h)=a_0+a_1h^2+a_2h^4$, with $a_0=0$; and, using (2.1.14) [where $\tilde{c}(0)=0$, $\tilde{c}(1)=\kappa_2(1)$, $\tilde{c}(2)=\kappa_2(2)$], we then obtain $a_1=\{16\kappa_2(1)-\kappa_2(2)\}/12$ and $a_2=\{\kappa_2(2)-4\kappa_2(1)\}/12$. So, (2.1.21) and (1.4.19) give

$$\kappa_2(h)/h^4=a_1/h^2+a_2-\{\sum_{i=1}^{h-2}(h-1-i)v_2(i)/(2+i)^{(3)}\}/h^3,$$
$$h=3,4,\ldots \tag{2.1.24}$$

which retrieves formula (4.12) of Cressie (1988). The modified Cressie polyvariogram of order 2 [for its definition, see (1.4.21)] is the third and final term on the right of (2.1.24), without the negative sign, whereas Cressie multiplies this by 6 to get his quadvariogram.

2.1.3 The legitimacy of the solution as a GCF

So far, we have found a $\kappa_b(h)$ given by (2.1.21), but can not yet claim that it is a generalised covariance function — since, up to now, we have only shown that this $\kappa_b(h)$ satisfies (1.4.9), whereas a GCF must satisfy (1.4.1) for all pairs of GIV of order b (not merely for the PIV) and with respect to all covariances (not just the variances). The next theorem establishes $\kappa_b(h)$ as a GCF.

In view of (1.4.1), the (i,k)-th element in K_{h+b} should be written as $\kappa_b(k-i)$. But, since we obtained $\kappa_b(j)$ as a solution of (1.4.9) under the symmetric assumption of (1.4.10),

$$\kappa_b(-j)=\kappa_b(j),\quad j=0,1,\ldots\,. \tag{2.1.25}$$

Theorem 2.1.3 *For integers b and $d, d\in\mathcal{N}, b\geq\max\{d-1,0\}$, and $\{Z(t)\}$ an $\tilde{I}_d$-series, $\kappa_b(j)$ is a GCF of order b [as (1.4.1) then holds for all $\boldsymbol{\lambda}\in\Lambda_b$ and $\boldsymbol{\mu}\in\Lambda_b$].*

Proof Due to (1.3.12), we can write

$$\boldsymbol{\lambda}=\sum_{i=0}^{m-b-1}\alpha_i\boldsymbol{\pi}_b^{(i,\ m-b-1-i)},\quad \boldsymbol{\mu}=\sum_{j=0}^{m-b-1}\beta_j\boldsymbol{\pi}_b^{(j,\ m-b-1-j)}. \tag{2.1.26a}$$

Also denote $(Z(m+t) \cdots Z(t))$ by $\mathbf{Z}'_m(t)$, so that

$$\begin{aligned}
&\mathrm{Cov}\{\mathbf{Z}'_m(0)\boldsymbol{\pi}_b^{(i,\, m-b-1-i)},\ \mathbf{Z}'_m(0)\boldsymbol{\pi}_b^{(j,\, m-b-1-j)}\} \\
= \;&\mathrm{Cov}\{\mathbf{Z}'_{b+1}(m-b-1-i)\boldsymbol{\pi}_b,\ \mathbf{Z}'_{b+1}(m-b-1-j)\boldsymbol{\pi}_b\} \\
&= \mathrm{Cov}\{\nabla^{b+1}Z(m-i),\ \nabla^{b+1}Z(m-j)\}. \qquad (2.1.26b)
\end{aligned}$$

Put

$$\eta(h) = \mathrm{Cov}\{\nabla^{b+1}Z(t),\ \nabla^{b+1}Z(t \pm h)\}, \qquad (2.1.27)$$

[where, when $b+1=d, \eta(h)=\gamma(h)$ given by (1.1.5).] Then, from (2.1.26), we have

$$\mathrm{Cov}\{\mathbf{Z}'_m(0)\boldsymbol{\lambda},\ \mathbf{Z}'_m(0)\boldsymbol{\mu}\} = \sum_{i=0}^{m-b-1-i}\ \sum_{j=0}^{m-b-1-j} \alpha_i\beta_j\eta(|\, i-j\,|). \qquad (2.1.28)$$

Now, from (1.4.10),

$$K_m = \begin{pmatrix} \kappa_b(0) & \cdots & \kappa_b(m) \\ \vdots & \ddots & \vdots \\ \kappa_b(-m) & \cdots & \kappa_b(0) \end{pmatrix} \qquad (2.1.29)$$

with $\kappa_b(j)$ given by (2.1.21) and (2.1.25). So, from (2.1.26a), we have

$$\boldsymbol{\lambda}'K_m\boldsymbol{\mu} = \sum_{i=0}^{m-b-1}\ \sum_{j=0}^{m-b-1} \alpha_i\beta_j\boldsymbol{\pi}'_b K_{j-i,b+1}\boldsymbol{\pi}_b, \qquad (2.1.30)$$

where

$$K_{h,b+1} = \begin{pmatrix} \kappa_b(h) & \cdots & \kappa_b(h+b+1) \\ \vdots & \ddots & \vdots \\ \kappa_b(h-b-1) & \cdots & \kappa_b(h) \end{pmatrix}. \qquad (2.1.31)$$

Then, in view of (2.1.28), (2.1.30) and K_m being symmetric, for proving that (1.4.1) holds, it is sufficient to show that

$$\eta(h) = \boldsymbol{\pi}'_b K_{h,b+1}\boldsymbol{\pi}_b, \quad h = 0, 1, \ldots\,. \qquad (2.1.32)$$

So, due to (2.1.4) and (2.1.21), it is equivalently sufficient to show that

$$\eta(h) = \boldsymbol{\pi}'_b \tilde{K}_{h,b+1}\boldsymbol{\pi}_b, \quad h = 0, 1, \ldots, \qquad (2.1.33)$$

where $\tilde{K}_{h,b+1}$ is the matrix as on the right of (2.1.31) with $\tilde{\kappa}_b(j)$ replacing $\kappa_b(j)$ everywhere.

On using (1.2.2), the right hand side of (2.1.33) becomes

$$(-1)^{b+1}\boldsymbol{\pi}_b'(\Delta^{b+1}\tilde{\kappa}_b(h)\cdots\ \Delta^{b+1}\tilde{\kappa}_b(h-b-1))' = (-1)^{b+1}\Delta^{2b+2}\tilde{\kappa}_b(h-b-1),$$

$$h = 0, 1, \ldots, \qquad (2.1.34)$$

and, with (2.3.6) below, this is seen to equal the left hand side of (2.1.33).

2.2 Variogram Formulae

We have defined two types of polyvariograms [see (1.4.20) and (1.4.21)]. For describing their properties and using them in model identification, it is important to obtain their formulae in terms of $\gamma(i)$, the ACVF of the hub series. As polyvariograms of each type can be expressed in terms of variograms $v_b(h)$ [refer to (1.4.19) and (1.4.13) for $\gamma_b^*(h)$], our first task is to obtain the formulae for $v_b(h)$. Then, the formulae for $\gamma_b(h)$ follow immediately. The formulae for $\gamma_b^*(h)$ will be discussed in the next section.

2.2.1 Integrated PIV

For a fixed $b \in \mathcal{N}$ and $h \in \mathcal{Z}^+$, $\{Y_b(h+b,t),\ t = 0,1,\ldots\}$ defined by (1.4.6) is stationary when $\{Z(t)\} \in \tilde{I}_d$ and

$$d \in \{0, 1, \ldots, b+1\} \qquad (2.2.1)$$

(see Theorem 1.3.4). So $v_b(h)$ in (1.4.8) can be defined by using any $t \in \mathcal{N}$, say $t = 0$. For simplicity of notation, abbreviate $Y_b(h+b,0)$ to $Y_b(h+b)$. Then, in view of (1.4.6),

$$Y_b(h+b) = \mathbf{Z}_{h+b}'(0)\boldsymbol{\pi}_b^h = \{\sum_{i=0}^{h+b} \pi_b^h(i)B^i\}Z(h+b). \qquad (2.2.2)$$

From (2.2.2), distinct expressions corresponding to different d can be derived; so we introduce the specific d as a second suffix, giving $Y_{bd}(h+b)$ [later on, we also write $v_{bd}(h)$, $\gamma_{bd}(h)$, $\gamma_{bd}^*(h)$ and so on].

Then, using (1.5.8), (2.2.2) gives

$$Y_{bd}(h+b) = \{\sum_{s=0}^{h-1} \frac{(b+s)^{(b)}}{b!} B^s(1-B)^{b+1}\}Z(h+b) \qquad (2.2.3)$$

$$= \{\sum_{s=0}^{h-1} \frac{(b+s)^{(b)}}{b!} B^s (1-B)^{b+1-d}\} W(h+b), \tag{2.2.4}$$

when $d \in \{0, 1, \ldots, b+1\}$ and $W(t)$ is the hub series. In particular, for $d = b+1$,

$$Y_{b,b+1}(h+b) = \{\sum_{s=0}^{h-1} \frac{(b+s)^{(b)}}{b!} B^s\} W(h+b). \tag{2.2.5}$$

Now, rewrite (2.2.4) as

$$Y_{bd}(h+b) = \{ \sum_{s=0}^{h+b-d} \pi_{bd}^h(s) B^s \} W(h+b) \tag{2.2.6}$$

and denote

$$\boldsymbol{\pi}_{bd}^h = (\pi_{bd}^h(0) \ \cdots \ \pi_{bd}^h(h+b-d))'. \tag{2.2.7}$$

Obviously,

$$\boldsymbol{\pi}_{b0}^h = \boldsymbol{\pi}_b^h \tag{2.2.8}$$

is a PIV, given by (1.5.15); and, from (2.2.5), we have

$$\pi_{b,b+1}^h(s) = (b+s)^{(b)}/b!, \quad s = 0, \ldots, h-1. \tag{2.2.9}$$

We next obtain the components of the general $\boldsymbol{\pi}_{bd}^h$.

Comparing (2.2.6) with (2.2.2), we get for $d \in \{0, 1, \ldots, b+1\}$,

$$(1-B)^d \sum_{s=0}^{h+b-d} \pi_{bd}^h(s) B^s = \sum_{s=0}^{h+b} \pi_{b0}^h(s) B^s. \tag{2.2.10}$$

Comparing the coefficients of B^s on both sides, we get [using (1.2.1)]

$$\nabla^d \pi_{bd}^h(s) = \pi_{b0}^h(s), \quad s = 0, \ldots, h+b; \quad d \in \{0, 1, \ldots, b+1\}; \tag{2.2.11}$$

with the boundary conditions

$$\pi_{bd}^h(s) = 0, \quad s < 0 \quad \text{or} \quad s > h+b-d. \tag{2.2.12}$$

Equivalently, (2.2.11) can be written as

$$\Delta^d \pi_{bd}^h(s-d) = \pi_{b0}^h(s); \tag{2.2.13}$$

and, using (1.2.4), we then obtain

$$\pi_{bd}^h(s-d) = \Sigma^d \pi_{b0}^h(s) = \frac{1}{(d-1)!} \sum_{j=0}^{s-d} (s-1-j)^{(d-1)} \pi_{b0}^h(j),$$

$$s = 0, \ldots, h + b. \quad d \in \{1, \ldots, b + 1\}. \tag{2.2.14}$$

Notice that $\pi_{bd}^h(s - d) = 0$ for $s < d$ or $s > h + b$, in agreement with (2.2.7).

As $\boldsymbol{\pi}_{b0}^h$ is a PIV, in view of (2.2.14), we call the $\boldsymbol{\pi}_{bd}^h$, $d = 1, \ldots, b+1$, the *integrated primary increment vectors* (IPIV). The IPIV are special and useful subsets of the GIV other than the PIV, and the following theorem extends Theorem 1.6.2.

Theorem 2.2.1 *The IPIV satisfy*

$$\boldsymbol{\pi}_{bd}^h \in \tilde{\Lambda}_{b-d}. \tag{2.2.15}$$

Proof From (2.2.10), $\nabla^d A^{-1}(\boldsymbol{\pi}_{bd}^h) = A^{-1}(\boldsymbol{\pi}_{b0}^h)$. So, looking at the proof of Theorem 1.6.2: $A^{-1}(\boldsymbol{\pi}_{bd}^h) = \nabla^{b-d+1} Q(B)$ where ∇ does not divide $Q(B)$, and (2.2.15) therefore holds. □

2.2.2 Explicit expressions for the IPIV

From (2.2.14), on using (1.5.15), we get the following extension to (1.5.15) for $d \in \{1, \ldots, b + 1\}$ and $s \in \{0, \ldots, h + b - d\}$

$$\begin{aligned}
\pi_{bd}^h(s) &= \tfrac{1}{(d-1)!} \textstyle\sum_{j=0}^{s} (s + d - 1 - j)^{(d-1)} \pi_{b0}^h(j) \\
&= \tfrac{1}{(d-1)!} \{(s + d - 1)^{(d-1)} + \textstyle\sum_{j=h}^{s} (s + d - 1 - j)^{(d-1)} \pi_{b0}^h(j)\} \\
&= \tfrac{1}{(d-1)!} \{(s + d - 1)^{(d-1)} + \textstyle\sum_{j=0}^{s-h} (s + d - 1 - h - j)^{(d-1)} \pi_{b0}^h(h + j)\} \\
&= \tfrac{1}{(d-1)!} \{(s + d - 1)^{(d-1)} - \tfrac{(h+b)^{(b+1)}}{b!} \textstyle\sum_{j=0}^{s-h} (-1)^j \begin{pmatrix} b \\ j \end{pmatrix} \tfrac{(s+d-1-h-j)^{(d-1)}}{h+j} \}.
\end{aligned} \tag{2.2.16}$$

So

$$\pi_{bd}^h(s) = (s + d - 1)^{(d-1)}/(d - 1)!\,, \quad s \in \{0, \ldots, h - 1\}, \tag{2.2.17}$$

and

$$\pi_{bd}^h(h+i) = \{(h+d-1+i)^{(d-1)} - \frac{(h + b)^{(b+1)}}{b!} \sum_{j=0}^{i} (-1)^j \begin{pmatrix} b \\ j \end{pmatrix} \frac{(d - 1 + i - j)^{(d-1)}}{h + j}\}/(d-1)!$$
$$i \in \{-1, \ldots, b - d\} \tag{2.2.18}$$

[where the uninteresting case of $i = -1$, which is the same as that of $s = h - 1$ in (2.2.17), is included to simplify subsequent argument].

Now (2.2.18) can be rewritten as

$$\pi^h_{bd}(h+i) = \frac{1}{b!}\sum_{j=i+1}^{b-d+1}(-1)^j \binom{b-d+1}{j}(h+b+i-j)^{(b)},$$

$$i = -1, \ldots, b-d, \quad d \in \{1, \ldots, b+1\}; \tag{2.2.19}$$

which is frequently a more convenient form, and which we now justify.

First, observe that (2.2.17) gives, for $s \in \{0, \ldots, h-1\}$,

$$\pi^h_{bd}(s) = \frac{1}{b!}\nabla^{b-d+1}(s+b)^b = \frac{1}{b!}\sum_{j=0}^{b-d+1}(-1)^j\binom{b-d+1}{j}(s+b-j)^b, \tag{2.2.20}$$

on using (1.2.1). Putting $s = h-1$, (2.2.20) shows that (2.2.19) holds for $i = -1$. For $i \in \{0, \ldots, b-d\}$, the justification that the right hand sides of (2.2.18) and (2.2.19) are equivalent is long; so, instead, we now verify these cases of (2.1.19) directly by induction.

Start by noting that, similarly to the way (2.2.11) was derived, we can also get

$$\pi^h_{b,d-1}(h+i) = \pi^h_{bd}(h+i) - \pi^h_{bd}(h+i-1), \quad i = 0, \ldots, b-d+1; \tag{2.2.21}$$

and, in particular, on putting $d = b+1$ and using (2.2.12) and (2.2.9),

$$\pi^h_{bb}(h) = \pi^h_{b,b+1}(h) - \pi^h_{b,b+1}(h-1) = -(h+b-1)^{(b)}/b!. \tag{2.2.22}$$

Now, looking at (2.2.19) for $i \in \{0, \ldots, b-d\}$ (giving no care when $d = b+1$), the inductive base $d = b$ is immediate, since (2.2.19) then reduces to (2.2.22). Next, suppose (2.2.19) is true for some $d \in \{2, \ldots, b\}$. Then, for $d-1$, (2.2.21) gives for $i = 0, \ldots, b-(d-1)$:

$$\begin{aligned}
\pi^h_{b,d-1}(h+i) &= \frac{1}{b!}\Big\{\sum_{j=i+1}^{b-d+1}(-1)^j\binom{b-d+1}{j}(h+b+i-j)^{(b)} \\
&\qquad - \sum_{j=i}^{b-d+1}(-1)^j\binom{b-d+1}{j}(h+b+i-1-j)^{(b)}\Big\} \\
&= \frac{1}{b!}\Big\{\sum_{j=i+1}^{b-d+2}(-1)^j\binom{b-d+1}{j}(h+b+i-j)^{(b)}
\end{aligned}$$

$$+\sum_{j=i+1}^{b-d+2}(-1)^j\begin{pmatrix} b-d+1 \\ j-1 \end{pmatrix}(h+b+i-j)^{(b)}\}$$

$$=\frac{1}{b!}\sum_{j=i+1}^{b-d+2}(-1)^j\begin{pmatrix} b-d+2 \\ j \end{pmatrix}(h+b+i-j)^{(b)},$$

so we have our inductive step. □

Combining (2.2.6), (2.2.17) and (2.2.19), we have the following theorem.

Theorem 2.2.2 *Suppose integer b and d satisfy $b+1 \geq d > 0$ and $\{Z(t)\} \in \tilde{I}_d$, then $Y_b(h+b)$ defined by (2.2.2) has the following formula*

$$Y_{bd}(h+b) = W(h+b) + \sum_{s=1}^{h-1}\{(s+d-1)^{(d-1)}/(d-1)!\}W(h+b-s)$$

$$+\frac{1}{b!}\sum_{i=0}^{b-d}\{\sum_{j=i+1}^{b-d+1}(-1)^j\begin{pmatrix} b-d+1 \\ j \end{pmatrix}(h+b+i-j)^{(b)}\}W(b-i), \quad h=1,2,\ldots; \quad (2.2.23a)$$

where $W(t) = \nabla^d Z(t)$.

Notice that, for the case $d = b+1$, the last summation of terms disappears. For the case of $d = 0$, (2.2.6) with (1.5.15) yields

$$Y_{b0}(h+b) = W(h+b) - \frac{(h+b)^{(b+1)}}{b!}\sum_{i=0}^{b}(-1)^i\begin{pmatrix} b \\ i \end{pmatrix}\frac{W(b-i)}{h+i}, \quad h=1,2,\ldots. \tag{2.2.23b}$$

2.2.3 Formulae for the variograms and γ_b-polyvariograms

Theorem 2.2.3 *For an $\tilde{I}_d$-series, $d \in \mathcal{N}$, when integer b satisfies (1.3.30), its variogram of order b can be written as*

$$v_{bd}(h) = E\{Y_{bd}^2(h+b)\} = \sum_{i=0}^{h+b-d} g_{bd}(h,i)\gamma(i), \quad h=1,2,\ldots, \tag{2.2.24}$$

where $\gamma(i)$ is the covariance function of the hub series $\{W(t)\}$ [see (1.1.5)] and

$$g_{bd}(h,i) = (2-\delta_i)\sum_{j=0}^{h+b-d-i}\pi_{bd}^h(i+j)\pi_{bd}^h(j), \quad i=0,\ldots,h+b-d, \tag{2.2.25}$$

where δ_i is the Kronecker delta function.

This theorem is a direct consequence of (1.4.8) and (2.2.6).

The most important case is $d = b + 1$, when (2.2.9) gives

$$g_{b,b+1}(h,i) = (2 - \delta_i)(\frac{1}{b!})^2 \sum_{j=0}^{h-1-i} (b+i+j)^{(b)}(b+j)^{(b)}, \quad i = 0, \ldots, h-1. \quad (2.2.26)$$

Clearly, we can extend the ranges of i in (2.2.25) and (2.2.26) to beyond $h + b - d$, giving zero $g_{bd}(h,i)$. So, in particular,

$$g_{b,b+1}(h,h) = 0. \quad (2.2.27)$$

Later on [see (2.4.3)], we will obtain an equivalent form of (2.2.26) which is easier for calculation and further analysis. For $d < b + 1$, no attempt has been made to get simple general explicit formulae as the $\pi_{bd}^h(h+i)$, $i = 0, \ldots, b-d$, have cumbersome expressions [see (2.2.18) and (2.2.19)]. However, we do give some explicit formulae for lower b — see the coefficients of $\gamma(i)$ in (2.2.32).

Theorem 2.2.4 *Under the same conditions as in the last theorem, for $d \in \{0, \ldots, b\}$,*

$$v_{bd}(h) = \sum_{i=1}^{h+b-d} g_{bd}(h,i)\dot{\gamma}(i), \quad h = 1, 2, \ldots, \quad (2.2.28)$$

where, $\dot{\gamma}(i) = \gamma(i) - \gamma(0)$.

Proof When $d \in \{0, \ldots, b\}$, from (2.2.4) and (2.2.6), it is clear that $1 - B$ is a factor of $\sum_{s=0}^{h+b-d} \pi_{bd}^h(s)B^s$, so

$$\sum_{s=0}^{h+b-d} \pi_{bd}^h(s) = 0, \quad (2.2.29)$$

and hence

$$\sum_{i=0}^{h+b-d} g_{bd}(h,i) = \{\sum_{s=0}^{h+b-d} \pi_{bd}^h(s)\}^2 = 0. \quad (2.2.30)$$

Then (2.2.28) follows from (2.2.24). □

One use of (2.2.30), or (2.2.28), is for checking appropriate explicit cases of (2.2.23), such as the ones in (2.2.32), below.

We now list some low-order formulae. From (2.2.23) and for $h = 1, 2, \ldots$:

$$
\begin{aligned}
Y_{00}(h) &= -W(0) + W(h) \\
Y_{01}(h) &= \textstyle\sum_{t=1}^{h} W(t) \\
Y_{10}(h+1) &= hW(0) - (h+1)W(1) + W(h+1) \\
Y_{11}(h+1) &= -hW(1) + \textstyle\sum_{t=2}^{h+1} W(t) \\
Y_{12}(h+1) &= \textstyle\sum_{t=2}^{h+1} (h+2-t)W(t) \\
Y_{20}(h+2) &= -(h+1)hW(0)/2 + (h+2)hW(1) \\
&\quad -(h+2)(h+1)W(2)/2 + W(h+2) \\
Y_{21}(h+2) &= (h+1)hW(1)/2 - (h+3)hW(2)/2 + \textstyle\sum_{t=3}^{h+2} W(t) \\
Y_{22}(h+2) &= -(h+1)hW(2)/2 + \textstyle\sum_{t=3}^{h+2} (h+3-t)W(t) \\
Y_{23}(h+2) &= (1/2) \textstyle\sum_{t=3}^{h+2} (h+4-t)^{(2)} W(t).
\end{aligned}
\tag{2.2.31}
$$

Next, from (2.2.24) and (2.2.25) [when $d = b + 1$, it is more convenient to use (2.4.3), below], we have for $h = 1, 2, \ldots$:

$$
\begin{aligned}
v_{00}(h) &= 2\{\gamma(0) - \gamma(h)\} \\
v_{01}(h) &= \textstyle\sum_{j=0}^{h-1} (2 - \delta_j)(h-j)\gamma(j) \\
v_{10}(h) &= 2\{(h^2 + h + 1)\gamma(0) - (h+1)h\gamma(1) - (h+1)\gamma(h) + h\gamma(h+1)\} \\
v_{11}(h) &= (h+1)h\gamma(0) - 2\textstyle\sum_{j=1}^{h} j\gamma(j) \\
v_{12}(h) &= (1/6)\textstyle\sum_{j=0}^{h-1} (2-\delta_j)(h-j+1)^{(2)}(2h+j+1)\gamma(j) \\
v_{20}(h) &= (1/2)[\{3(h+2)(h+1)^2 h + 4\}\gamma(0) - 4(h+2)(h+1)^2 h\gamma(1) \\
&\quad +(h+2)(h+1)^2 h\gamma(2) - 2\{(h+2)(h+1)\gamma(h) \\
&\quad -2(h+2)h\gamma(h+1) + (h+1)h\gamma(h+2)\}] \\
v_{21}(h) &= (1/2)[(h+2)(h+1)^2 h\gamma(0) - \{(h+2)(h+1)^2 h + 4\}\gamma(1) \\
&\quad -4\textstyle\sum_{j=2}^{h} j\gamma(j) + 2(h+1)h\gamma(h+1)]
\end{aligned}
\tag{2.2.32}
$$

$$v_{22}(h) = (1/12)[(h+2)^{(3)}(3h+1)\gamma(0)$$
$$-4\sum_{j=1}^{h}\{(h+2)^{(3)}-(j+1)^{(3)}\}\gamma(j)]$$
$$v_{23}(h) = (1/120)\sum_{j=0}^{h-1}(2-\delta_j)(h-j+2)^{(3)}\{3h(2h+j+4)$$
$$+(j+2)(j+1)\}\gamma(j).$$

In the formulae for $v_{10}(h)$, when $h = 1$, the coefficients of $\gamma(1)$ and $\gamma(h)$ should be added together [the sum being $g_{10}(1,1)$]. The interpretation for $v_{20}(h)$ is similar, when $h = 1$ or $h = 2$. When $d \in \{0,\ldots,b\}$, we can rewrite the formulae of (2.2.32) in the form (2.2.28).

From (1.4.20) and (2.2.24), under the conditions of Theorem 2.2.2, the polyvariogram can be written as

$$\gamma_{bd}(h) = \sum_{i=0}^{h+b-d} g_{bd}(h,i)\gamma(i)/h^{2b}, \quad h = 1,2,\ldots, \tag{2.2.33}$$

or from (1.4.20) and (2.2.28), when $d < b+1$,

$$\gamma_{bd}(h) = \sum_{i=1}^{h+b-d} g_{bd}(h,i)\dot{\gamma}(i)/h^{2b}, \quad h = 1,2,\ldots . \tag{2.2.34}$$

2.3 Formulae for Primary Generalized Covariance Functions

2.3.1 Equation linking the GCF and ACVF

Whether $d = b+1$ or $d < b+1$, we can always write (2.2.3) as

$$Y_b(h+b) = \{\sum_{s=0}^{h-1}\pi_{b,b+1}^{h}(s)B^s\}\nabla^{b+1}Z(h+b), \quad \pi_{b,b+1}^{h}(s) = \frac{(b+s)^{(b)}}{b!}. \tag{2.3.1}$$

Instead of (2.2.24), we have

$$v_b(h) = E\{Y_b^2(h+b)\} = \sum_{i=0}^{h-1} g_{b,b+1}(h,i)\eta(i), \quad h = 1,2,\ldots, \tag{2.3.2}$$

where $\eta(i)$ is defined by (2.1.27) and $g_{b,b+1}(h,i)$ is defined by (2.2.26).

A particular solution of (1.4.9), $\tilde{\kappa}_b(h)$ $(h = b+1, b+2,\ldots)$, was given by (1.4.19). In fact, the range of (1.4.19) can be extended to $h = 0,1,\ldots$; as (1.4.19) then correctly gives $\tilde{\kappa}_b(h) = 0$ for $h = 0, \ldots, b$, since we always interpret a summation as zero when it's lower limit exceeds its upper limit. Like (2.1.25), put

$$\tilde{\kappa}_b(-h) = \tilde{\kappa}_b(h), \quad h = 0, 1, \ldots, \tag{2.3.3}$$

so,

$$\tilde{\kappa}_b(h) = 0, \quad h = -b, \ldots, b. \tag{2.3.4}$$

According to Theorem 2.1.3, $\tilde{\kappa}_b(h)$ is a special GCF achieved by putting $c_b(h) \equiv 0$ in (2.1.21), and we call it the *primary generalized covariance function* (PGCF). However, the proof of Theorem 2.1.3 still requires that (2.3.6) be established — which we now do.

Employing (1.4.17) and the product formula (1.2.16), and then (1.4.16) followed by (1.4.13), and finally (2.3.2) and (2.2.27), we get for $s = 0, 1, \ldots$:

$$\begin{aligned}
\Delta^{b+1}\tilde{\kappa}_b(s) &= \Delta^{b+1}\{s\tilde{\xi}_b(s)\} = s\{\Delta\Delta^b\tilde{\xi}_b(s)\} + (b+1)\Delta^b\tilde{\xi}_b(s+1) \\
&= (s+b+1)\Delta^b\tilde{\xi}_b(s+1) - s\Delta^b\tilde{\xi}_b(s) = (-1)^{b+1}b!\Delta v_b(s)/\{2(s+b)^{(b)}\} \\
&= \frac{(-1)^{b+1}b!}{2(s+b)^{(b)}}\sum_{i=0}^{s}\{g_{b,b+1}(s+1,i) - g_{b,b+1}(s,i)\}\eta(i).
\end{aligned}$$

Then, by (2.2.26) we have (for all integers $s \geq 0$)

$$\Delta^{b+1}\tilde{\kappa}_b(s) = \frac{(-1)^{b+1}}{b!2}\sum_{i=0}^{s}(2-\delta_i)(s+b-i)^{(b)}\eta(i).$$

In view of (2.3.4), $\Delta^{b+1}\tilde{\kappa}_b(s) = 0$, $s = -b, \ldots, -1$. Now, for $h = 1, 2, \ldots$, put $t = h - b - 1$ (so $t = -b, -b+1, \ldots$). Then (1.2.2) gives

$$\begin{aligned}
\Delta^{2b+2}\tilde{\kappa}_b(t) &= (-1)^{b+1}\sum_{k=0}^{b+1}(-1)^k\binom{b+1}{k}\Delta^{b+1}\tilde{\kappa}_b(t+k) \\
&= \frac{1}{b!2}\sum_{k=0}^{b+1}\{(-1)^k\binom{b+1}{k}\sum_{i=0}^{t+k}(2-\delta_i)(t+k+b-i)^{(b)}\eta(i)\} \\
&= \frac{1}{b!}\{(\sum_{i=0}^{t}\sum_{k=0}^{b+1} + \sum_{i=t+1}^{t+b}\sum_{k=i-t}^{b+1})(1-\frac{\delta_i}{2})(-1)^k\binom{b+1}{k}(t+k+b-i)^{(b)}\eta(i)\} \\
&\quad +(1-\frac{\delta_{t+b+1}}{2})(-1)^{b+1}\eta(t+b+1).
\end{aligned} \tag{2.3.5}$$

Notice that, when $k = 0, \ldots, i-t-1$ and $i = t+1, \ldots, t+b$, then $0 \le t+k+b-i < b$ and so $(t+k+b-i)^{(b)} = 0$; and, hence, the lower limit $i-t$ of k in the final summation can be replaced by 0. Thus the whole summation part on the right of (2.3.5) is zero because $\nabla^{b+1}(t+k+b-i)^{(b)} = 0$. So, for $h = 1, 2, \ldots$,

$$\Delta^{2b+2}\tilde{\kappa}_b(h-b-1) = (-1)^{b+1}\eta(h), \tag{2.3.6}$$

which also holds for $h = 0$. This is because (2.1.27), followed by (1.4.9), a comparison between (2.1.29) and (2.1.31), and finally (2.1.21) with (2.1.4), yields

$$\eta(0) = \mathrm{Var}\{\nabla^{b+1}Z(t)\} = \boldsymbol{\pi}_b' K_{1+b}\boldsymbol{\pi}_b = \boldsymbol{\pi}_b' K_{0,b+1}\boldsymbol{\pi}_b = \boldsymbol{\pi}_b'\tilde{K}_{0,b+1}\boldsymbol{\pi}_b;$$

which, on using (2.1.34), gives $\eta(0) = (-1)^{b+1}\Delta^{2b+2}\tilde{K}_b(-b-1)$.

since $\Delta = B^{-1}\nabla$, (2.3.6) can also be written as

$$\Delta^{b+1}\nabla^{b+1}\tilde{\kappa}_b(h) = (-1)^{b+1}\eta(h), \quad h = 0, 1, \ldots\,. \tag{2.3.7}$$

Since the general solution $\kappa_b(h)$ of (1.4.9) differs from $\tilde{\kappa}_b(h)$ by $c_b(h)$, a polynomial of degree $2b$ [see (2.1.21) and (2.1.13)], $\kappa_b(h)$ also satisfies (2.3.6) and (2.3.7). However that is immaterial for our present purpose, as the modified Cressie polyvariograms, $\gamma_b^*(h)$ [defined by (1.4.21)], depend only on $\tilde{\kappa}_b(h)$.

Given $b \in \mathcal{N}$, suppose $\{Z(t)\} \in \tilde{I}_d$ with $d \in \{0, 1, \ldots, b+1\}$, then (2.1.27) gives $\eta(h) = \mathrm{Cov}\{\nabla^{b-d+1}W(t), \nabla^{b-d+1}W(t \pm h)\}$, and it is easy to verify that

$$\eta(h) = (-1)^{b-d+1}\Delta^{b-d+1}\nabla^{b-d+1}\gamma(h), \quad h = 0, 1, \ldots\,. \tag{2.3.8}$$

From (2.3.8), we see that both equations (2.3.6) and (2.3.7) directly link the GCF, $\tilde{\kappa}_b(h)$, and $\gamma(h)$, the ACVF of the hub series. So, $\tilde{\kappa}_b(h)$ may be expressed in terms of the hub autocovariances [as a particular solution of (2.3.6) or (2.3.7)].

2.3.2 The PGCF as a particular solution of (2.3.6)

We now consider the particular solution of (2.3.6) satisfying the conditions (2.3.3) and (2.3.4) [in fact, the only use of (2.3.3) is $\tilde{\kappa}_b(-b-1) = \tilde{\kappa}_b(b+1)$]. Suppose $\{Z(t)\}$ is an $\tilde{I}_d$-series, where d satisfies (2.2.1). Again, to distinguish the distinct expressions for $\tilde{\kappa}_b(h)$ corresponding to different d, we introduce the specific d as a second suffix, giving $\tilde{\kappa}_{bd}(h)$. Put

$$\zeta(h) = \Delta^d \nabla^d \tilde{\kappa}_{bd}(h), \quad h = -(b-d+1),\ -(b-d), \dots . \tag{2.3.9}$$

[For simplicity of notation we omit the subscripts b and d from $\zeta(h)$ and, also, from $e(h)$ introduced in (2.3.12), below.] Combining (2.3.7) and (2.3.8), we get

$$\Delta^{b-d+1} \nabla^{b-d+1} \zeta(h) = \Delta^{b-d+1} \nabla^{b-d+1} (-1)^d \gamma(h), \quad h = 0, 1, \dots . \tag{2.3.10}$$

When $\mathbf{d = b + 1}$, (2.3.10) simply gives

$$\zeta(h) = (-1)^d \gamma(h), \quad h = 0, 1, \dots . \tag{2.3.11}$$

When $\mathbf{0 \le d < b + 1}$, as $\Delta^{b+1-d} \nabla^{b+1-d}$ annihilates any polynomial of degree $2b - 2d + 1$, say,

$$e(h) = e_0 + e_1 h + \ldots + e_{2b-2d+1} h^{2b-2d+1}. \tag{2.3.12}$$

[$e(h)$ may be written in other convenient forms such as a "polynomial" with factorial powers or the forms favoured later in this subsection — see (2.3.28) and (2.3.30).] So, (2.3.10) gives

$$\zeta(h) = e(h) + (-1)^d \gamma(h), \quad h = -(b-d+1), -(b-d), \dots . \tag{2.3.13}$$

There are $2b - 2d + 2$ constants in $e(h)$ which need to be determined. In view of (2.3.4) and (2.3.9),

$$\zeta(h) = 0, \quad h = -(b-d), \dots, b-d. \tag{2.3.14}$$

This provides $2b - 2d + 1$ initial conditions for $\zeta(h)$ and we see that, due to (2.3.4), all elements in $\tilde{K}_{0,b+1}$ [given by (2.1.31) with $\tilde{\kappa}_{bd}(j)$ replacing $\kappa_{bd}(j)$ everywhere] are zero except two, those in the lower-left and upper-right corners. In view of (2.1.33), (2.3.3) and (2.3.8), these two are

$$\tilde{\kappa}_{bd}\{\pm(b+1)\} = (-1)^{b+1} \eta(0)/2 = (-1)^d \{\Delta^{b-d+1} \nabla^{b-d+1} \gamma(h)\}_{h=0}/2. \tag{2.3.15}$$

We get the final initial condition for $\zeta(h)$ from (2.3.9), (2.3.4) and (2.3.15), viz

$$\zeta\{-(b-d+1)\} = \tilde{\kappa}_{bd}\{-(b+1)\} = (-1)^d \{\Delta^{b-d+1} \nabla^{b-d+1} \gamma(h)\}_{h=0}/2. \tag{2.3.16}$$

Then the $2b-2d+2$ constants in $\varepsilon(h)$ can be determined from the $(2b-2d+1)+1$ conditions in (2.3.14) and (2.3.16); and, hence, $\zeta(h)$ in (2.3.13) may be obtained.

Having got $\zeta(h)$, consider the solution of $\tilde{\kappa}_{bd}(h)$ in (2.3.9), or equivalently, the solution of

$$\Delta^{2d}\tilde{\kappa}_{bd}(h) = \zeta(h+d), \quad h = -(b+1), -b, \ldots . \tag{2.3.17}$$

When $\mathbf{d = 0}$, (2.3.17) and (2.3.13) simply give

$$\tilde{\kappa}_{bd}(h) = \zeta(h) = \varepsilon(h) + \gamma(h), \quad h = -(b+1), -b, \ldots . \tag{2.3.18}$$

When $\mathbf{0 < d \leq b+1}$, in view of (1.2.4) (with r most conveniently chosen as d) and (2.1.3), the general solution of (2.3.17) is: for $h = -(b+1), -b, \ldots,$

$$\begin{aligned}\tilde{\kappa}_{bd}(h) &= \alpha(h) + \frac{1}{(2d-1)!}\sum_{i=d}^{h-2d}(h-1-i)^{(2d-1)}\zeta(i+d)\\ &= \alpha(h) + \frac{1}{(2d-1)!}\sum_{i=0}^{h-d}(h+d-1-i)^{(2d-1)}\zeta(i),\end{aligned} \tag{2.3.19}$$

where $\alpha(h)$ is a polynomial of degree $2d-1$.

Note that all the terms in this last summation are also polynomials in h of degree $2d-1$. So, if an $r > d$ had been chosen, then those terms lost from the summation would have been added into $\alpha(h)$, to give $\alpha^*(h)$ say; while, if $r < d$ were the choice, the extra terms in the summation would have been subtracted from $\alpha(h)$, giving a different $\alpha^*(h)$. In all such cases, $\alpha^*(h)$ will still be a polynomial of degree $2d-1$, and analysis can proceed [although, unlike (2.3.21) below, the formula for $\alpha^*(h)$ will include hub-series autocovariances as well as $\gamma(0)$].

There are $2d$ constants in $\alpha(h)$ which need to be determined. In view of (2.3.4), and noticing that the summation in (2.3.19) disappears for $h = -b, \ldots, d-1$, we see that $\alpha(h) = \tilde{\kappa}_{bd}(h) = 0$ in those cases. So, when $0 < d \leq b$, $\alpha(h)$ has $b+d \geq 2d$ zeros; while, when $d = b+1$, $\alpha(h)$ has $2d-1$ zeros. But $\alpha(h)$ is a polynomial of degree $2d-1$, so

$$\alpha(h) = 0 \quad (0 < d \leq b) \tag{2.3.20}$$

and

$$\alpha(h) = ah(h^2-1)\cdots\{h^2-(d-1)^2\} = a(h+d-1)^{(2d-1)} \quad (d = b+1),$$

where a is a constant which we now obtain. Putting $h = -d$

$$\alpha(-d) = -(2d-1)!a,$$

while, from (2.3.19) followed by (2.3.15),

$$\alpha(-d) = \tilde{\kappa}_{bd}\{-(b+1)\} = (-1)^d\gamma(0)/2.$$

Togather these give a, and hence:

$$\alpha(h) = \frac{(-1)^{d-1}\gamma(0)}{2(2d-1)!}(h+d-1)^{(2d-1)} \quad (d = b+1). \tag{2.3.21}$$

The following theorem summarizes the above.

Theorem 2.3.1 *Given $b \in \mathcal{N}$, for an $\tilde{I}_d$-series with d satisfying (2.2.1), the PGCF, $\tilde{\kappa}_{bd}(h)$, defined by (1.4.19), (2.3.3) and (2.3.4), satisfies the difference equation (2.3.6) [or equivalently, (2.3.7)]. The unique solution of the equation has the form of (2.3.18) when $d = 0$; or (2.3.19) when $0 < d \le b+1$. In (2.3.19), when $0 < d < b+1$, $\alpha(h) = 0$; when $d = b+1$, $\alpha(h)$ is given by (2.3.21). When $d = b+1$, the $\zeta(i)$ in (2.3.19) are given by (2.3.11); while, when $0 \le d \le b$, the $\zeta(i)$ in (2.3.18) and (2.3.19) are given by (2.3.13). In (2.3.13), $e(h)$ is a polynomial of degree $2b-2d+1$, whose $2b-2d+2$ constants can be obtained from (2.3.14) and (2.3.16).*

As far as $\zeta(h)$ is concerned, the simplest case is $d = b+1$. In view of (2.3.11), (2.3.19) and (2.3.21), we have

$$\tilde{\kappa}_{b,b+1}(h) = \frac{(-1)^{b+1}}{(2b+1)!}\sum_{i=0}^{h-b-1}(1-\frac{\delta_i}{2})(h+b-i)^{(2b+1)}\gamma(i). \tag{2.3.22}$$

For all other cases, although $\alpha(h)$ disappears in (2.3.19), (2.3.13) shows that $\zeta(h)$ contains a polynomial $e(h)$ which needs to be determined. The larger the difference $b-d$ is, the higher the degree of $e(h)$.

Now, from(2.3.13) and (2.3.14),

$$(-1)^d e(i) = -\gamma(i), \quad i = d-b, \ldots, b-d; \tag{2.3.23a}$$

while, from (2.3.13) and (2.3.16),

$$(-1)^d e(d-b-1) = -\gamma(d-b-1) + \{\Delta^{b-d-1}\nabla^{b-d-1}\gamma(h)\}_{h=0}/2. \tag{2.3.23b}$$

Then the $2d-2b+2$ particular values of $e^*(h) = (-1)^d e(h)$, given by (2.3.23), are sufficient to determine $e^*(h)$ uniquely when $b-d$ is fixed.

As an example, when $b = d$, (2.3.23) gives [for $e(h) = e_0 + e_1 h$] $(-1)^d e(h) = -\gamma(0)$ $(h = 0, -1)$ — from which $e(h) \equiv (-1)^{d-1}\gamma(0)$. Then, if $b = 0$, (2.3.18) gives $\tilde{\kappa}_{00}(h) = -\gamma(0) + \gamma(h)$; while, if $b > 0$, (2.3.19) gives [on using (2.3.20) and (2.3.13)]:

$$\tilde{\kappa}_{bb}(h) = \frac{1}{(2b-1)!}\sum_{i=0}^{h-b}(h+b-1-i)^{(2b-1)}\{(-1)^{b+1}\gamma(0) + (-1)^b\gamma(i)\}$$

$$= \frac{(-1)^b}{(2b-1)!}\sum_{i=0}^{h-b}\{1 - \frac{(h+b)}{2b}\delta_i\}(h+b-1-i)^{(2b-1)}\gamma(i). \qquad (2.3.24)$$

However, in general, it seems easier to obtain $e(h)$ in the following way. For $b \geq d$, denote

$$f(h) = h^2(h^2-1)\cdots\{h^2-(b-d)^2\} = h(h+b-d)^{(2b-2d+1)} \qquad (2.3.25)$$

and

$$f_i(h) = \begin{cases} f(h)/h^2, & i = 0, \\ f(h)/(h-i), & i \in \{-(b-d),\ldots,-1\} \cup \{1,\ldots,b-d\}. \end{cases} \qquad (2.3.26)$$

Put

$$e(h) = \sum_{i=-(b-d)}^{b-d} \beta_i f_i(h), \qquad (2.3.27)$$

which is again a polynomial of degree $2b-2d+1$, but notice that (2.3.27) is not an exactly equivalent form of (2.3.12) since there are only $2b-2d+1$ free constants β_i, rather than $2b-2d+2$. However, (2.3.27) is legitimate if we can determine β_i such that (2.3.14) and (2.3.16) are satisfied. Using this form, (2.3.13) and (2.3.14) immediately give

$$\beta_i = (-1)^{d+1}\gamma(i)/f_i(i), \quad i = -(b-d),\ldots,b-d.$$

Then we get

$$e(h) = (-1)^{d+1}\sum_{i=-(b-d)}^{b-d}\{f_i(h)/f_i(i)\}\gamma(i). \qquad (2.3.28)$$

Further, put $\tilde{f}_0(h) = f_0(h)$, and (when $d < b$) put

$$\tilde{f}_i(h) = f(h)/(h^2-i^2), \quad i = 1,\ldots,b-d. \qquad (2.3.29)$$

Then, $-f_{-i}(i) = f_i(i) = 2i\tilde{f}_i(i)$ and (2.3.28) becomes

$$e(h) = (-1)^{d+1}\sum_{i=0}^{b-d}\{\tilde{f}_i(h)/\tilde{f}_i(i)\}\gamma(i) \tag{2.3.30}$$

which is easier to apply than (2.3.28).

Theorem 2.3.2 *When $0 \le d \le b$, the polynomial $e(h), h = -(b-d+1), -(b-d), \ldots$, [which appears in (2.3.13)] is given by (2.3.28) or its equivalent form (2.3.30), which shows that $e(h)$ has only even powers and depends only on $b-d$ [except for a factor $(-1)^{d+1}$].*

Proof We derived (2.3.28) under condition (2.3.14), so only condition (2.3.16) needs to be verified, or equivalently [noting (2.3.13)], the following relation may be checked (where $c = b - d$)

$$e(-c-1) = -(-1)^d\{(-1)^c\binom{2c+2}{c+1}\frac{\gamma(0)}{2} + (-1)^{c-1}\binom{2c+2}{c}\gamma(1)$$

$$+\cdots+\binom{2c+2}{1}\gamma(c)\}. \tag{2.3.31}$$

In view of (2.3.26),

$$f_0(c+1) = (2c+1)!/(c+1)$$

and

$$f_0(0) = (-1)^c(c!)^2.$$

So,

$$\frac{f_0(c+1)}{f_0} = (-1)^c\binom{2c+1}{c+1}. \tag{2.3.32}$$

While, for $i \in \{-c, \ldots, -1\} \cup \{1, \ldots, c\}$,

$$f_i(c+1) = (c+1)\{(2c+1)!\}/(c+1-i)$$

and

$$f_i(i) = (c+i)!(c-i)!(-1)^{c-i}i.$$

So

$$\frac{f_i(c+1)}{f_i(i)} = \frac{(-1)^{c-i}(2c+2)!}{(c+i)!(c+1-i)!2i};$$

and, then,

$$\frac{f_i(c+1)}{f_i(i)} + \frac{f_{-i}(c+1)}{f_{-i}(-i)} = (-1)^{c-i}\begin{pmatrix} 2c+2 \\ c+1-i \end{pmatrix}.$$

Then using this and (2.3.32) in (2.3.28) gives (2.3.31). □

For $d = 0$, we may get $\tilde{\kappa}_{b0}(h)$ directly from (2.3.18) and (2.3.30). For example, when $b = 1$, we have $b - d = 1, f(h) = h^2(h^2 - 1), \tilde{f}_0(0) = f_0(0) = (h^2 - 1)_{h=0} = -1$, and $\tilde{f}_1(1) = (h^2)_{h=1} = 1$. So, $\tilde{\kappa}_{10}(h) = (-1)\{-(h^2 - 1)\gamma(0) + h^2\gamma(1)\} + \gamma(h)$. But, for $0 < d < b + 1$, in view of (2.3.13) and (2.3.20), (2.3.19) becomes

$$\tilde{\kappa}_{bd}(h) = \frac{1}{(2d-1)!}\sum_{i=0}^{h-d}(h + d - 1 - i)^{(2d-1)}\{e(i) + (-1)^d\gamma(i)\}$$

$$h = -(b+1), -b, \ldots$$

We will list all the formulae for $\tilde{\kappa}_{bd}(h)$ when $b = 0, 1$ and 2 in the next subsection.

2.3.3 Formulae for the PGCFs and γ_b^*-Polyvariograms

The first means of obtaining formulae for $\tilde{\kappa}_{bd}(h)$ is to use (1.4.18) (when $b = 0$) or (1.4.19) (when $b > 0$), first, and then (1.4.13) and (2.2.24). For $b > 0$, from these relations we have

$$\tilde{\kappa}_{bd}(h) = \frac{h}{(b-1)!}\sum_{j=1}^{h-b}(h - 1 - j)^{(b-1)}\frac{b!(-1)^{b+1}}{2(b+j)^{(b+1)}}\sum_{i=0}^{j+b-d} g_{bd}(j,i)\gamma(i) \qquad (2.3.33)$$

$$= (-1)^{b+1}\frac{bh}{2}\{\sum_{i=0}^{1+b-d}\sum_{j=1}^{h-b} + \sum_{i=2+b-d}^{h-d}\sum_{j=i-b+d}^{h-b}\}\frac{(h-1-j)^{(b-1)}}{(b+j)^{(b+1)}}g_{bd}(j,i)\gamma(i).$$

For $i = 0, \ldots, h - d$, writing

$$g_{bd}^*(h,i) = (-1)^{b+1}\frac{bh}{2}\sum_{j=max(1,i-b+d)}^{h-b}\frac{(h-1-j)^{(b-1)}}{(b+j)^{(b+1)}}g_{bd}(j,i), \qquad (2.3.34)$$

we get

$$\tilde{\kappa}_{bd}(h) = \sum_{i=0}^{h-d} g_{bd}^*(h,i)\gamma(i), \quad h = b+1, b+2, \ldots, \qquad (2.3.35)$$

and $\tilde{\kappa}_{bd}(h) = 0,\ h = 0, \ldots, b.$

For $b = 0$, if we still put $\tilde{\kappa}_{bd}(h)$ as above, then (1.4.18) implies

$$g_{0d}^*(h,i) = -g_{0d}(h,i)/2, \quad h = 1, 2, \ldots. \qquad (2.3.36)$$

When $d \in \{0, \ldots, b\}$, using (2.3.33) and (2.3.35) followed by (2.2.30) gives

$$\sum_{i=0}^{h-d} g_{bd}^{*}(h,i) = (-1)^{b+1} \frac{bh}{2} \sum_{j=1}^{h-b} \{ \frac{(h-1-j)^{(b-1)}}{(b+j)^{(b+1)}} \sum_{k=0}^{j+b-d} g_{bd}(j,k) \} = 0, \tag{2.3.37}$$

which can be used for checking particular explicit cases of (2.3.35), like (2.3.39) below.

Corresponding to Theorems 2.2.2 and 2.2.3 we then have the following theorem:

Theorem 2.3.3 *For an $\tilde{I}_d$-series and integers b and d satisfying (1.3.30), $\tilde{\kappa}_{bd}(h)$, the PGCF given by (1.4.19), (2.2.3) and (2.2.4), can be expressed for $h \geq 0$ by (2.3.35) with (2.3.34) (when $b > 0$) or (2.3.36) (when $b = 0$), where the $g_{bd}(j,i)$ are given by (2.2.25). For $d \in \{0, \ldots, b\}$,*

$$\tilde{\kappa}_{bd}(h) = \sum_{i=1}^{h-d} g_{bd}^{*}(h,i)\dot{\gamma}(i), \quad h = b+1, b+2, \ldots. \tag{2.3.38}$$

where $\dot{\gamma}(i) = \gamma(i) - \gamma(0)$.

Obtaining the numerical values of $g_{bd}^{*}(h,i)$, via (2.3.37) and (2.2.25), is straightforward in principle and easily programmed for specific cases. But, to achieve fully simplified formulae, more and more algebraic manipulation is needed as b increases. Even for $b = 2$, considerable effort is required. [Although (partially) automating the work, by using a computer-algebra package such as Maple, would greatly help with higher b.] However, explicit expressions for the PGCFs may be obtained more directly [without the need to formally calculate the $g_{bd}(j,k)$ and $g_{bd}^{*}(h,i)$], by using Theorems 2.3.1 and 2.3.2. The results, for the nine possible cases when $b = 0$, 1 and 2, are listed in (2.3.39). Comparing these formula with (2.2.32), we see that these PGCFs are, in general, rather simpler than the corresponding variograms. We may consider all these formulae for $h = 0, 1, \ldots$ [when $h = 0, \ldots, b$, $\tilde{\kappa}_{bd}(h) = 0$; and the formulae are also available for $h = -(b+1), -b, \ldots$, due to (2.3.3) and (2.3.4)]. The coefficients of the $\gamma(i)$ in these formulae are in fact, just the $g^*_{bd}(h,i)$ given by (2.3.34).

$$\begin{aligned}
\tilde{\kappa}_{00}(h) &= -\gamma(0)+\gamma(h) \\
\tilde{\kappa}_{01}(h) &= -\{\textstyle\sum_{i=0}^{h-1}(2-\delta_i)(h-i)\gamma(i)\}/2 \\
\tilde{\kappa}_{10}(h) &= (h^2-1)\gamma(0)-h^2\gamma(1)+\gamma(h) \\
\tilde{\kappa}_{11}(h) &= h^{(2)}\gamma(0)/2-\textstyle\sum_{i=1}^{h-1}(h-i)\gamma(i) \\
\tilde{\kappa}_{12}(h) &= \{\textstyle\sum_{i=0}^{h-2}(2-\delta_i)(h+1-i)^{(3)}\gamma(i)\}/12 \\
\tilde{\kappa}_{20}(h) &= -(h^2-1)(h^2-4)\gamma(0)/4+h^2(h^2-4)\gamma(1)/3 \\
&\quad -h^2(h^2-1)\gamma(2)/12+\gamma(h) \\
\tilde{\kappa}_{21}(h) &= -h^{(3)}(h+3)\gamma(0)/12+(h-1)^{(2)}(h^2+3h+6)\gamma(1)/12 \\
&\quad -\textstyle\sum_{i=2}^{h-1}(h-i)\gamma(i) \\
\tilde{\kappa}_{22}(h) &= -(h+1)^{(4)}\gamma(0)/24+\{\textstyle\sum_{i=1}^{h-2}(h+1-i)^{(3)}\gamma(i)\}/6 \\
\tilde{\kappa}_{23}(h) &= -\{\textstyle\sum_{i=0}^{h-3}(2-\delta_i)(h+2-i)^{(5)}\gamma(i)\}/240.
\end{aligned} \tag{2.3.39}$$

From (1.4.21) and (2.3.35), under the conditions of Theorem 2.3.3, the modified Cressie polyvariogram has the form

$$\gamma^*_{bd}(h)=(-1)^{b+1}\sum_{i=0}^{h-d} g^*_{bd}(h,i)\gamma(i)/h^{2b}, \quad h=b+1,b+2,\ldots, \tag{2.3.40}$$

or from (1.4.21) and (2.4.38), when $d<b+1$,

$$\gamma^*_{bd}(h)=(-1)^{b+1}\sum_{i=1}^{h-d} g^*_{bd}(h,i)\dot{\gamma}(i)/h^{2b}, \quad h=b+1,b+2,\ldots\ . \tag{2.3.41}$$

2.4 Polyvariogram Asymptotes

For an $\tilde{I}_d$-series, if an integer b satisfies (1.3.30), then both polyvariograms of order b, $\gamma_b(h)$ and $\gamma^*_b(h)$, exist. This section will reveal a striking contrast which plays a key role in deriving certain procedures for identifying d : when $d=b+1, \gamma_b(h)$ and $\gamma^*_b(h)$ have linear increasing trends their asymptotes are straight lines with positive slopes; but when $d \leq b, \gamma_b(h)$ and $\gamma_b^*(h)$ level out — have horizontal lines as their asymptotes. This feature was first discovered by Cressie (1988) for his semivariogram ($b = 0$), linvariogram ($b = 1$) and quadvariogram ($b = 2$). He also derived the explicit formulae for their asymptotes under the assumption that the hub series, $\{W(t)\}$, is white noise [$\gamma(i) = 0$ for all $i \neq 0$]. We deal with general b and general stationary $\{W(t)\}$.

2.4.1 Two identities

Identity 1 With the forward differencing operator operating on j,

$$(r+j)^{(b)}(s+j)^{(b)} = \Delta\{\sum_{k=0}^{b}(-1)^k b^{(k)}(r+j)^{(b-k)}(s+k+j)^{(b+1+k)}/(b+1+k)^{(k+1)}\}. \quad (2.4.1)$$

Proof By repeatedly using (1.2.10) and (1.2.17), we have for (2.4.1):

$$\begin{aligned}
\text{LHS} &= (r+j)^{(b)}\{\Delta(s+j)^{(b+1)}\}/(b+1) \\
&= [\Delta\{(r+j)^{(b)}(s+j)^{(b+1)}\} - \{\Delta(r+j)^{(b)}\}(s+1+j)^{(b+1)}]/(b+1) \\
&= [\Delta\{(r+j)^{(b)}(s+j)^{(b+1)}\} - b(r+j)^{(b-1)}\{\Delta(s+1+j)^{(b+2)}/(b+2)\}]/(b+1) \\
&= \Delta\{\sum_{k=0}^{1}(-1)^k b^{(k)}(r+j)^{(b-k)}(s+k+j)^{(b+1+k)}/(b+1+k)^{(k+1)}\} \\
&\quad +b^{(2)}(r+j)^{(b-2)}\{\Delta(s+2+j)^{(b+3)}\}/(b+3)^{(3)} \\
&\;\vdots \\
&= \Delta\{\sum_{k=0}^{b-1}(-1)^k b^{(k)}(r+j)^{(b-k)}(s+k+j)^{(b+1+k)}/(b+1+k)^{(k+1)}\} \\
&\quad +(-1)^b b^{(b)}(r+j)^{(0)}\{\Delta(s+b+j)^{(2b+1)}\}/(2b+1)^{(b+1)} = \text{RHS}. \quad \square
\end{aligned}$$

Identity 2

$$\sum_{k=0}^{b}\{(-1)^k b^{(k)}/(b+1+k)^{(k+1)}\} = \frac{1}{2b+1} \quad (2.4.2)$$

Proof

$$\text{LHS} = \sum_{k=0}^{b}\{(-1)^k \frac{b!}{(b-k)!}\frac{b!}{(b+1+k)!}\frac{(2b+1)!}{(2b+1)!}\}$$

$$= \frac{(b!)^2}{(2b+1)!}\sum_{k=0}^{b}(-1)^k\left\{\binom{2b}{b-k}+\binom{2b}{b-k-1}\right\}$$

$$= \frac{(b!)^2}{(2b+1)!}\binom{2b}{b} = \text{RHS}. \quad \square$$

2.4.2 Asymptotes of the $\gamma_{bd}(h)$ (case $d = b+1$)

Lemma 2.4.1 *The $g_{b,b+1}(h,i)$ defined by (2.2.26) can be expressed as*

$$g_{b,b+1}(h,i) = (2-\delta_i)(\frac{1}{b!})^2\sum_{k=0}^{b}(-1)^k b^{(k)}\frac{(b+h)^{(b-k)}(b+k+h-i)^{(b+1+k)}}{(b+1+k)^{(k+1)}} \tag{2.4.3}$$

and

$$0 < g_{b,b+1}(h,i) \le (2-\delta_i)(h+2b)^{2b+1}/(b!)^2. \quad 0 \le i \le h-1, \ h \ge 1. \tag{2.4.4}$$

Further, given any η $(0 \le \eta < 1)$ and $c > 0$, then

$$g_{b,b+1}(h,i) = (2-\delta_i)h^{2b+1}/\{(2b+1)(b!)^2\} + O(h^{2b+\eta}), \quad 0 \le i \le ch^\eta. \tag{2.4.5}$$

Proof Using (2.4.1) (with r and s replaced by $b+i$ and b, respectively) on the right of (2.2.26), we get:

$$\begin{aligned}&\{(b!)^2/(2-\delta_i)\}g_{b,b+1}(h,i)\\ &= \sum_{j=0}^{h-1-i}\Delta\{\sum_{k=0}^{b}(-1)^k b^{(k)}(b+i+j)^{(b-k)}(b+k+j)^{(b+1+k)}/(b+1+k)^{(k+1)}\}\\ &= \sum_{k=0}^{b}(-1)^k b^{(k)}(b+h)^{(b-k)}(b+k+h-i)^{(b+1+k)}/(b+1+k)^{(k+1)} - 0,\end{aligned}$$

which proves (2.4.3).

Now, for $b \ge 0$, $0 \le k \le b$, and $0 \le i \le h-1$, it is certainly true that

$$(b+h)^{(b-h)}(b+k+h-i)^{(b+1+k)} \le (2b+h)^{2b+1},$$

and

$$\sum_{k=0}^{b}\left|\frac{(-1)^k b^{(k)}}{(b+1+k)^{(k+1)}}\right| = \frac{1}{b+1}+\frac{b}{(b+2)(b+1)}+\ldots+\frac{b!}{(2b+1)\ldots(b+1)} \le 1.$$

So (2.4.3) gives the right inequality of (2.4.4), while the left inequality is immediate from (2.2.26).

Next, for all i such that $0 \leq i \leq ch^{\eta}$,

$$(b+k+h)^{(b+1+k)} \geq (b+k+h-i)^{(b+1+k)} \geq \{b+k+h(1-ch^{\eta-1})\}^{(b+1+k)}. \quad (2.4.6)$$

The LHS of (2.4.6) is $h^{b+1+k} + O(h^{b+k})$, while its RHS is $h^{b+1+k} + O(h^{b+\eta+k})$. Thus we have

$$(b+h)^{(b-k)}(b+k+h-i)^{(b+1+k)} = h^{2b+1} + O(h^{2b+\eta}). \quad (2.4.7)$$

Then (2.4.5) follows directly from (2.4.3) on using (2.4.2) and (2.4.7). □

Next consider $\{Z(t)\} \in \tilde{I}_{b+1}$ and suppose that the autocovariances for the hub series, $\{W(t) = \nabla^{b+1}Z(t)\}$, satisfy

$$\sum_{h=0}^{\infty} | \gamma(h) | < \infty. \quad (2.4.8)$$

Then, certainly, the spectral density, $s(\omega)$, for $\{W(t)\}$ exists and can be expressed as

$$s(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-\iota\omega h}\gamma(h), \quad \omega \in [-\pi, \pi]. \quad (2.4.9)$$

However, we are going to need a rather stronger condition than (2.4.8), viz: for some $c > 0$, there exist a $c^* > 0$ and a certain η $(0 \leq \eta < 1)$, such that

$$\sum_{i > h(\eta)} | \gamma(i) | \leq ch^{\eta-1}, \quad \forall\, h \in \mathcal{Z}^+, \quad (2.4.10)$$

where $h(\eta)$ is the integer part of c^*h^{η}. In the case with $\eta = 0$, (2.4.10) implies that the ACVF for $\{W(t)\}$ is zero for $i > h(0)$, i.e. the hub series is a white noise or moving average series. If $\{W(t)\}$ is a proper ARMA series, (2.4.10) holds for any $\eta \in (0,1)$, since $| \gamma(i) |$ then decreases geometrically fast. When η tends to 1, condition (2.4.10) tends to (2.4.8). (2.4.8) is a sufficient condition for the absence of the deterministic component from the Wold decomposition of $\{W(t)\}$ [see (3.1.6), below].

Theorem 2.4.1 *Suppose $\{Z(t)\} \in \tilde{I}_{b+1}$ and (2.4.10) holds. Then, as $h \to \infty$,*

$$v_{b,b+1}(h) = 2\pi s(0)h^{2b+1}/\{(2b+1)(b!)^2\} + O(h^{2b+\eta}). \quad (2.4.11)$$

Proof In view of (2.2.24), we may put

$$v_{b,b+1}(h) = \{\sum_{i=0}^{h(\eta)} + \sum_{i=h(\eta)+1}^{h-1}\} g_{b,b+1}(h,i)\gamma(i). \qquad (2.4.12)$$

Now, due to (2.4.4) and (2.4.10),

$$\Big| \sum_{i=h(\eta)+1}^{h-1} g_{b,b+1}(h,i)\gamma(i) \Big| \le \max_{i>h(\eta)} g_{b,b+1}(h,i) \sum_{i=h(\eta)+1}^{\infty} |\gamma(i)| \le c'h^{2b+\eta} \qquad (2.4.13a)$$

where $c' > 0$ is a constant; while, from (2.4.5) and (2.4.8),

$$\sum_{i=0}^{h(\eta)} g_{b,b+1}(h,i)\gamma(i) = \frac{h^{2b+1}}{(2b+1)(b!)^2} \sum_{i=-h(\eta)}^{h(\eta)} \gamma(i) + O(h^{2b+\eta}). \qquad (2.4.13b)$$

Also (2.4.9) and (2.4.10) give,

$$\Big| \sum_{i=-h(\eta)}^{h(\eta)} \gamma(i) - 2\pi s(0) \Big| \le 2 \sum_{i>h(\eta)} |\gamma(i)| \le 2ch^{\eta-1}. \qquad (2.4.14)$$

Then (2.4.11) follows from (2.4.12) through (2.4.14). □

From (2.4.11) we see that, when $\eta = 0$ (white noise or moving average cases), the order of the remainder is the smallest. When $\{W(t)\}$ is a strongly dependent series, the order of the remainder approaches that of the leading term. Generally, $s(0) > 0$. However, $s(0) = 0$ is possible [although (1.3.24) must hold]; in such a situation (2.4.11) is meaningless.

Corollary 2.4.1 *Suppose* $\{Z(t)\} \in \tilde{I}_{b+1}$, *(2.4.10) holds and* $s(0) > 0$. *Then, as* $h \to \infty$, *the polyvariogram* $\gamma_{b,b+1}(h) = v_{b,b+1}(h)/h^{2b}$ *can be expressed as*

$$\gamma_{b,b+1}(h) = 2\pi s(0)h/\{(2b+1)(b!)^2\} + O(h^{\eta}). \qquad (2.4.15)$$

It is then tempting to conclude that $\gamma_{b,b+1}(h)$ has an asymptote which is a fixed straight line with positive slope, $2\pi s(0)/\{(2b+1)(b!)^2\}$; but this is not always so. As an example, consider [see(2.2.32)]

$$\gamma_{01}(h) = v_{01}(h) = h\{\gamma(0) + 2\sum_{j=1}^{h-1}\gamma(j)\} - 2\sum_{j=1}^{h-1} j\gamma(j).$$

Obviously $\gamma(0) + 2\sum_{j=1}^{h-1}\gamma(j) \to 2\pi s(0)$, as $h \to \infty$; and, if $-2\sum_{j=1}^{h-1} j\gamma(j)$ converges, then $\gamma_{01}(h)$ has the fixed line with slope $2\pi s(0)$ and intercept $-2\sum_{j=1}^{\infty} j\gamma(j)$ as its asymptote. But, if the summation diverges or oscillates, then there would not be a fixed finite intercept. However this is immaterial as $\gamma_{01}(h)$ still has an upward trend which is roughly linear with slope of around $2\pi s(0)$, when h is large.

2.4.3 Asymptotes of the $\gamma^*_{bd}(h)$ (case $d = b+1$)

We have, from (2.3.22), that for $h \geq b+1$:

$$g^*_{b,b+1}(h,i) = (1-\frac{\delta_i}{2})\frac{(-1)^{b+i}}{(2b+1)!}(h+b-i)^{(2b+1)}, \quad 0 \leq i \leq h-b-1. \qquad (2.4.16)$$

Analogous to (2.4.4) and (2.4.5), we can similarly show that

$$0 < (-1)^{b+1} g^*_{b,b+1}(h,i) \leq (2-\delta_i)(h+b)^{2b+1}/\{(2b+1)!2\},$$

$$0 \leq i \leq h-b-1, \quad h \geq b+1, \qquad (2.4.17)$$

and

$$g^*_{b,b+1}(h,i) = (1-\frac{\delta_i}{2})(-1)^{b+1}h^{2b+1}/(2b+1)! + O(h^{2b+\eta})$$

$$0 \leq i \leq ch^{\eta}, \quad 0 \leq \eta < 1. \qquad (2.4.18)$$

Then the proof of the following theorem is similar to that of Theorem 2.4.1.

Theorem 2.4.2 *Suppose $\{Z(t)\} \in \tilde{I}_{b+1}$ and (2.4.10) holds. Then, as $h \to \infty$,*

$$\tilde{\kappa}_{b,b+1}(h) = (-1)^{b+1}\pi s(0)h^{2b+1}/(2b+1)! + O(h^{2b+\eta}). \qquad (2.4.19)$$

Corollary 2.4.2 *Suppose $\{Z(t)\} \in \tilde{I}_{b+1}$, (2.4.10) holds and $s(0) > 0$. Then, as $h \to \infty$, the polyvariogram $\gamma^*_{b,b+1}(h) = (-1)^{b+1}\tilde{\kappa}_{b,b+1}(h)/h^{2b}$ can be expressed as*

$$\gamma^*_{b,b+1}(h) = \pi s(0)h/(2b+1)! + O(h^{\eta}). \qquad (2.4.20)$$

2.4.4 Asymptotes of the $\gamma_{bd}(h)$ and $\gamma^*_{bd}(h)$ (cases $d < b+1$)

From (2.2.32) and (2.3.39) we see that, when $b \in \{0,1,2\}$ and $d < b+1$, $v_{bd}(h)$ and $\tilde{\kappa}_{bd}(h)$ are of order $O(h^{2b})$. But now, rather than get the general explicit formulae of the asymptotes via the $g_{bd}(h)$ and $g^*_{bd}(h)$, it is more convenient to use the representation of an $\tilde{I}_d$-series discussed in the next chapter. The proof of the following theorem is therefore delayed until subsection 3.2.3, below.

Theorem 2.4.3 *Suppose $b \in Z^+$, $\{Z(t)\} \in \tilde{I}_d$ $(0 \leq d \leq b)$, and (2.4.8) holds. Then, as $h \to \infty$,*

$$v_{bd}(h) = h^{2b}\{\Delta^{b-d}\nabla^{b-d}\gamma(h)\}_{h=0}/(b!)^2 + O(h^{2b-1}). \qquad (2.4.21)$$

$$\tilde{\kappa}_{bd}(h) = h^{2b}(-1)^{b+1}\{\Delta^{b-d}\nabla^{b-d}\gamma(h)\}_{h=0}/(2b)! + O(h^{2b-1}), \tag{2.4.22}$$

where

$$\{\Delta^c\nabla^c\gamma(h)\}_{h=0} = \binom{2c}{c}\gamma(0) + 2\sum_{i=1}^{c}(-1)^i\binom{2c}{c+i}\gamma(i). \tag{2.4.23}$$

The asymptotes of $v_{00}(h)$ and $\tilde{\kappa}_{00}(h)$ are horizontal lines with intercepts of $2\gamma(0)$ and $-\gamma(0)$, respectively.

Notice that the $v_{00}(h)$ intercept does not agree with (2.4.21) extended to the case of $b = 0$, but the $\tilde{\kappa}_{00}(h)$ intercept agrees with (2.4.22) similarly extended. All the other special cases in (2.2.32) and (2.3.39) agree with the theorem, when the $\gamma(j)$ have suitable convergence rates. From this theorem and the definitions (1.4.20) and (1.4.21), we immediately get the following corollary.

Corollary 2.4.3 *Under the same conditions as in Theorem 2.4.3 (extended to include $b = 0$), the asymptotes of $\gamma_{bd}(h)$ and $\gamma^*_{bd}(h)$ are horizontal lines with intercepts of*

$$(1 + \delta_b\delta_d)\{\Delta^{b-d}\nabla^{b-d}\gamma(h)\}_{h=0}/(b!)^2 \quad \text{and} \quad \{\Delta^{b-d}\nabla^{b-d}\gamma(h)\}_{h=0}/(2b)!\ , \tag{2.4.24}$$

respectively. The remainders are of order $O(h^{-1})$.

2.4.5 A discussion

For spatial processes that are intrinsic random fuctions (see subsection 1.3.1) of order b and $b-1$, respectively, Matheron (1973, pp 451-454; also cited by Cressie, 1988) proved a pair of results which for univariate time series reduce to the following. For an I_{b+1}-series, as $h \to \infty$,

$$\kappa_b(h)/h^{2b+2} \to 0. \tag{2.4.25}$$

For an I_b-series,

$$|\ \kappa_b(h)\ | \le a_1 + a_2|\ h\ |^{2b}, \quad a_1 > 0,\ a_2 > 0. \tag{2.4.26}$$

Comparing (2.4.25) and (2.4.26) with the results in this chapter, we see that a substantial improvement to this aspect of Matheron's work has been achieved for the time series case. Although (2.4.25) and (2.4.26) hint at distinct behaviours for the $\kappa_b(h)$ in the two cases of $d = b+1$ and $d < b+1$, they are insufficient even for setting up a graphical contrast. With our specially defined $\tilde{\kappa}_b(h)$ [or $v_b(h)$], the graphical features are clear and we may also develop formal inferential procedures for determining d (see Chapter 4).

Chapter 3

Series Structures

3.1 The Representation and the Decomposition

3.1.1 The basic theorem

In subsection 1.3.1, we introduced the transformation of an intrinsic random function (IRF), $Z(\mathbf{x})$, of order b:

$$Z^*(\lambda) = \int Z(\mathbf{x})\lambda(d\mathbf{x}), \quad \lambda \in \Lambda_b. \tag{3.1.1}$$

Consider any IRF, $Y(\mathbf{x})$, that only differs from $Z(\mathbf{x})$ by some polynomial in the elements of $\mathbf{x}$ of degree not greater than b and with coefficients that can be random variables. Now by (1.3.1), for each $\lambda \in \Lambda_b$, $\lambda(d\mathbf{x})$ annihilates all such polynomials. So, then, $Y(\mathbf{x})$ will have the same transform as $Z(\mathbf{x})$, that is

$$Z^*(\lambda) = \int Y(\mathbf{x})\lambda(d\mathbf{x}), \quad \forall\ \lambda \in \Lambda_b. \tag{3.1.2}$$

Matheron (1973) called a $Y(\mathbf{x})$ satisfying (3.1.2) a representation of $Z(\mathbf{x})$ of order b.

Back to the time series case, suppose that $\{Z(t)\} \in \tilde{I}_d$ and integers b and d satisfy (1.3.30). Then we may define $Y_b(h+b,t)$ by (1.4.6) which, for a fixed h, gives a stationary series as t varies. Take $t = 0$ and again, as in Chapter 2, abbreviate $Y_b(h+b,0)$ to $Y_b(h+b)$. Then, due to (1.5.15), (2.2.2) gives

$$Y_b(h+b) = Z(h+b) + \frac{1}{b!}\sum_{i=0}^{b}(-1)^{i+1}\binom{b}{i}\frac{(h+b)^{(b+1)}}{h+i}Z(b-i),$$
$$h = 1,2,\ldots, \tag{3.1.3}$$

which has the form $Y_b(h+b) = Z(h+b) +$ a polynomial of degree b in h. Thus, $Y_b(h+b)$ is a representation of $Z(h+b)$ of order b. We call (3.1.3) the *primary representation of order b*; and, from (1.4.6) and (1.4.8), we see that *the variogram of order b is the variance of the primary representation of order b as h varies.*

In general, given $b \in \mathcal{N}$, we may express $Y_{bd}(h+b)$ as (2.2.4), for any $d \in \{0, \ldots, b+1\}$, but the most important case is $d = b+1$ (no overdifferencing), that is (2.2.5). This case also implies $d > 0$. When we have $d \in \mathcal{Z}^+$ and $\{Z(t)\} \in \tilde{I}_d$, we emphasise that the representation of order $b = d-1$ is being used, by rewriting (2.2.5) as:

$$Y_{d-1,d}(h+d-1) = \frac{1}{(d-1)!}\sum_{j=0}^{h-1}(d-1+j)^{(d-1)}W(h+d-1-j),$$
$$h = 1, 2, \ldots, \qquad (3.1.4)$$

where $W(t) = \nabla^d Z(t)$. Then, putting $t = h + d - 1$, we get

$$Y_{d-1,d}(t) = \frac{1}{(d-1)!}\sum_{j=0}^{t-d}(d-1+j)^{(d-1)}W(t-j),$$
$$t = d, d+1, \ldots \; . \qquad (3.1.5)$$

Let us recall some basic theory in second-order stationary time series. If $\{W(t)\}$ is such a series with zero mean, then $\{W(t)\}$ has the Wold decomposition (see, say, Doob 1953; Hannan 1970)

$$W(t) = U(t) + V(t). \qquad (3.1.6)$$

$U(t)$ is the regular component which can be expressed as a general linear series

$$U(t) = \sum_{j=0}^{\infty}\alpha_j A(t-j), \quad \sum_{j=0}^{\infty}\alpha_j^2 < \infty; \qquad (3.1.7)$$

where $\{A(t)\}$ is white noise such that

$$E\{A(t)\} = 0, \quad E\{A(s)A(t)\} = \delta_{s-t}\sigma^2 \qquad (3.1.8)$$

with δ_i again the Kronecker delta function. $V(t)$ is the deterministic component which can be expressed as

$$V(t) = \sum_{r=-\infty}^{\infty}\xi_r e^{i\omega_r t}, \quad \omega_{-r} = \omega_r \in [-\pi, \pi] \quad (\omega_r \neq 0, \text{ when } r \neq 0). \qquad (3.1.9)$$

If the complex random variables, ξ_r, satisfy

$$\xi_0 \equiv 0, \quad \xi_{-r} = \overline{\xi}_r, \quad E(\xi_r) = 0, \quad E(\xi_r \overline{\xi}_q) = s_r \delta_{r-q}, \tag{3.1.10}$$

where $\overline{\xi}$ denotes the complex conjugate of ξ and $s_r > 0$ is defined by (3.1.13), below, then $V(t)$ is real and $E\{V(t)\} = 0$.

In fact, if we use the spectral representation, we may put

$$W(t) = \int_{-\pi}^{\pi} e^{i\omega t}\xi(d\omega) = \int_{-\pi}^{\pi} e^{i\omega t}\xi^{(U)}(d\omega) + \int_{-\pi}^{\pi} e^{i\omega t}\xi^{(V)}(d\omega). \tag{3.1.11}$$

Then

$$s(\omega) = E\{\mid \xi^{(U)}(d\omega) \mid^2\}/d\omega = \frac{\sigma^2}{2\pi} \mid \sum_{j=0}^{\infty} \alpha_j e^{-i\omega j} \mid^2 \tag{3.1.12}$$

is the spectral (power) density of $\{U(t)\}$ and

$$s_r = E\{\mid \xi^{(V)}(d\omega_r) \mid\}^2\} = E\{\mid \xi_r \mid^2\} \tag{3.1.13}$$

is the power of the sinusoid $\xi_r e^{i\omega_r t}$ – a jump at ω_r in the (power) spectrum of $\{W(t)\}$.

Later on, when we talk about a representation of order b, b can be any integer satisfying $b \geq \max\{d-1, 0\}$; but, when we talk about the decomposition, we usually refer to the case of $b = d - 1$.

Theorem 3.1.1 *For any $d \in \mathcal{Z}^+$, an $\tilde{I}_d$-series, $\{Z(t)\}$, having hub series, $\{W(t)\}$, can be decomposed into three components, viz:*

$$Z(t) = Y_{d-1,d}(t) + R(t) = P(t) + Q(t) + R(t), \quad t = d, d+1, \ldots, \tag{3.1.14}$$

where

$$P(t) = \frac{1}{(d-1)!}\sum_{j=0}^{t-d}(d-1+j)^{(d-1)}U(t-j), \tag{3.1.15}$$

$$Q(t) = \frac{1}{(d-1)!}\sum_{r=-\infty}^{\infty} \xi_r\{\sum_{j=0}^{t-d}(d-1+j)^{(d-1)}e^{i\omega_r(t-j)}\}, \tag{3.1.16}$$

$$R(t) = \frac{1}{(d-1)!}\sum_{j=0}^{d-1}(-1)^{d-1-j}\binom{d-1}{j}\frac{t^{(d)}}{t-j}Z(j), \tag{3.1.17}$$

and $U(t)$, which can be expressed as (3.1.7), is the regular component of $W(t)$. $\{U(t)\}$ has spectral density (3.1.12) which satisfies (1.3.24). ω_r and ξ_r satisfy (3.1.9) and (3.1.10), which characterize the deterministic component, $V(t)$, of $W(t)$.

Proof The result follows directly from (3.1.3), (3.1.5), (3.1.6) and (3.1.9) [in (3.1.3), b should be replaced by $d-1$ and $h+b$ should be replaced by t]. Since $\{Z(t)\} \in \tilde{I}_d$, $\{W(t)\} \in \tilde{I}_0$ and hence (1.3.24) holds. □

$P(t)$ and $Q(t)$ provide the d-times repeated summations of the two Wold decomposition parts of the hub series, which are evidently anticipated for an integrated stationary time series. $R(t)$ is a rather less obvious component dependent only on the d starting values of $\{Z(t)\}$. We call $P(t)$, $Q(t)$ and $R(t)$ the regular, deterministic and initial components, respectively, of $Z(t)$. As we shall see (subsection 3.2.1, below), $P(t)$ is the component which dominates the divergence rate of the whole series. Both $Q(t)$ and $R(t)$ are completely predictable; but, although $R(t)$ is evidently a polynomial of degree $d-1$, the functional nature of $Q(t)$ is not immediately clear. However, the next theorem shows that $Q(t)$ is, in fact, a sum of sinusoids and polynomials of degree $d-1$. So we conclude that, *for an integrated stationary time series of degree d, the only possible trends hidden in it are polynomials (with degree not exceeding $d-1$) and sinusoids.*

3.1.2 The deterministic component

Lemma 3.1.1 *In (3.1.16), replace t by $h+d-1$ and denote the coefficient of $\xi_r e^{i\omega_r t}$ by $q_d(h,\omega_r)$, that is*

$$q_d(h,\omega) = \sum_{j=0}^{h-1}(d-1+j)^{(d-1)}e^{-i\omega j}; \tag{3.1.18}$$

and, for simplicity of notation, put

$$e(\omega) = 1 - e^{-i\omega}, \quad \omega \neq 0. \tag{3.1.19}$$

Then

$$q_d(h,\omega) = (d-1)!e^{-d}(\omega)\{1 - e^{-i\omega h}\sum_{j=0}^{d-1} e^{j}(\omega)(h-1+j)^{(j)}/j!\}. \tag{3.1.20}$$

Proof For simplicity of notation, denote $q_d(h,\omega)$ in (3.1.18) by q_d [for fixed h and ω], then

$$q_{d+1} = \sum_{j=0}^{h-1}(d+j)^{(d)}e^{-i\omega j} = \sum_{j=0}^{h-1}(d+j)(d-1+j)^{(d-1)}e^{-i\omega j}$$

$$= \quad dq_d + \sum_{j=1}^{h}(d-1+j)^{(d)}e^{-i\omega j} - (d-1+h)^{(d)}e^{-i\omega h}$$

$$= \quad dq_d + e^{-i\omega}q_{d+1} - (h+d-1)^{(d)}e^{-i\omega h},$$

or

$$q_{d+1} - e^{-1}(\omega)dq_d = -e^{-1}(\omega)(h+d-1)^{(d)}e^{-i\omega h}. \tag{3.1.21}$$

To solve (3.1.21), first put

$$q_d = (d-1)!e^{-d}(\omega)p_d, \quad d = 1,2\dots . \tag{3.1.22}$$

Then (3.1.21) becomes

$$\Delta p_d = -e^{d}(\omega)(h+d-1)^{(d)}e^{-i\omega h}/d!, \quad d = 1,2,\dots,$$

which has the general solution

$$p_d = c - e^{-i\omega h}\sum_{j=1}^{d-1} e^{j}(\omega)(h+j-1)^{(j)}/j!, \quad d = 1,2;\dots, \tag{3.1.23}$$

where the constant, c, can be determined by equating the right hand sides of (3.1.22) [using (3.1.23)] and (3.1.18), when $d - 1$, giving

$$c = e(\omega)\sum_{j=0}^{h-1} e^{-i\omega j} = 1 - e^{-i\omega h}. \tag{3.1.24}$$

Then (3.1.20) follows from (3.1.22), (3.1.23) and (3.1.24). □

Theorem 3.1.2 *Under the same conditions as for Theorem 3.1.1,*

$$Q(t) = \sum_{r=-\infty}^{\infty} \frac{\xi_r}{(1-e^{-i\omega_r})^{d}}e^{i\omega_r t} - \sum_{k=0}^{d-1}\{\sum_{j=k}^{d-1}(-1)^{j}s(j,k)\sum_{r=-\infty}^{\infty}\frac{\xi_r e^{i\omega_r(d-1)}}{(1-e^{-i\omega_r})^{d-j}}\}(d-1-t)^{k},$$

$$t = d, d+1, \dots; \tag{3.1.25}$$

where ω_r and ξ_r are given by (3.1.9) and (3.1.10), and the $s(j,k)$ are Stirling numbers of the first kind (see, say, Bayer 1987, p 527) defined by

$$t^{(j)} = \sum_{k=0}^{j} s(j,k)t^{k}, \quad t \in Z. \tag{3.1.26}$$

Proof First note that, for any integer $h > 0$, using (3.1.26), we have

$$(h-1+j)^{(j)} = (-1)^j(-h)^{(j)} = (-1)^j \sum_{k=0}^{j} s(j,k)(-h)^k. \tag{3.1.27}$$

Next, since $h - 1 = t - d$, Lemma 3.1.1 followed by (3.1.27) gives:

$$\begin{aligned}
&\frac{1}{(d-1)!}\sum_{j=0}^{t-d}(d-1+j)^{(d-1)}e^{i\omega_r(t-j)} \\
=\ & e^{-d}(\omega_r)e^{i\omega_r t}\{1 - e^{-i\omega_r(t-d+1)}\sum_{j=0}^{d-1} e^j(\omega_r)(h-1+j)^{(j)}/j!\} \\
=\ & e^{-d}(\omega_r)[e^{i\omega_r t} - e^{i\omega_r(d-1)}\sum_{j=0}^{d-1}\{-e(\omega_r)\}^j \sum_{k=0}^{j} s(j,k)(d-1-t)^k]. \\
=\ & e^{-d}(\omega_r)[e^{i\omega_r t} - e^{i\omega_r(d-1)}\sum_{k=0}^{d-1}\{\sum_{j=k}^{d-1}(-e(\omega_r))^j s(j,k)\}(d-1-t)^k].
\end{aligned}$$

Then (3.1.25) follows, on using (3.1.16) and (3.1.19). □

In the RHS of (3.1.25), the first summation (r runs from $-\infty$ to ∞) is a sum of sinusoids and the second summation (k runs from 0 to $d-1$) is a polynomial. Both the sinusoids and the polynomial have stochastic coefficients. Notice that, for (3.1.25) to be meaningful, it is sufficient that the series $\sum_{r=-\infty}^{\infty} \xi_r/(1-e^{-i\omega_r})^d$ should converge in some probabilistic sense. [In most practical situations, there are only a finite number of ω_r. But, theoretically, there could be a denumerable infinite number of ω_r, when the summations over r in (3.1.25) might diverge if there is a subsequence of $\{\omega_r\}$ that converges to 0.]

3.1.3 The primary representation

Matheron (1973) defined a special type of representation which, when restricted to $\mathcal{R}$, is characterized as follows.

Suppose $\{\lambda_j : j = 0,\ldots,b\}$ is a set of measures with the same finite support, $\{0,\ldots,b\}$, such that

$$\int t^i \lambda_j(dt) = \delta_{i-j}, \quad i,j = 0,\ldots,b, \tag{3.1.28}$$

where δ_i is the Kronecker delta function. (The corresponding vectors will have all their elements zero after the first $b+1$ elements.) Denote the Dirac delta function by $\delta_{\{t\}}$ — that is, for any continuous function $f(s)$,

$$\int f(s)\delta_{\{t\}}(ds) = f(t). \tag{3.1.29}$$

(A vector corresponding to $\delta_{\{t\}}$ has all its elements zero except the $(t+1)$-th which is 1). Let

$$\pi_{\{t\}} = \delta_{\{t\}} - \sum_{j=0}^{b} t^j \lambda_j, \tag{3.1.30}$$

then

$$\int s^i \pi_{\{t\}}(ds) = t^i - \sum_{j=0}^{b} t^j \int s^i \lambda_j(ds) = 0, \quad i = 0, \ldots, b.$$

So, in view of the measure analogue of (1.3.1) [see (1.3.1) with $n = 1$], the vector corresponding to $\pi_{\{t\}}$ is in Λ_b, and we may also write

$$\pi_{\{t\}} \in \Lambda_b. \tag{3.1.31}$$

Now, for $\{Z(t)\} \in \tilde{I}_d$, $d \le b+1$, and using (3.1.30) followed by (3.1.29) and (3.1.2),

$$Y(t) \stackrel{\text{def}}{=} \int Z(s)\pi_{\{t\}}(ds) = Z(t) - \sum_{j=0}^{b} t^j Z^*(\lambda_j), \quad t = 0, 1, \ldots . \tag{3.1.32}$$

Then $Y(t)$ is a representation of $Z(t)$ of order b since any λ, $\lambda \in \Lambda_b$, annihilates t^j $(j = 0, \ldots, b)$ and, hence, $\int Y(t)\lambda(dt) = \int Z(t)\lambda(dt) = Z^*(\lambda)$. Moreover, in view of (3.1.1), on using first (3.1.32) and then (3.1.28), we see that $Y(t)$ has the property:

$$Y^*(\lambda_j) = \int Y(t)\lambda_j(dt) = Z^*(\lambda_j) - \sum_{i=0}^{b} \{\int t^i \lambda_j(dt)\} Z^*(\lambda_i) = 0,$$

$$j = 0, \ldots, b. \tag{3.1.33}$$

Since all the λ_j have the same support, $\{0, \ldots, b\}$, we can write

$$\lambda_j(dt) = \{\boldsymbol{\delta}(dt)\}'\boldsymbol{\mu}_j, \quad j = 0, \ldots, b, \tag{3.1.34}$$

where $\{\boldsymbol{\delta}(dt)\}' = (\delta_{\{b\}}(dt) \ \ldots \ \delta_{\{0\}}(dt))$ and $\boldsymbol{\mu}_j = (\mu_{bj} \ \ldots \ \mu_{0j})'$. Then (3.1.28) and (3.1.29) give

$$I_{b+1} = V_b(\boldsymbol{\mu}_b \ \ldots \ \boldsymbol{\mu}_0) \ .$$

Here I_{b+1} is the unit matrix and V_b is the Vandermonde matrix

$$V_b = \begin{pmatrix} b^b & \cdots & 1^b & 0 \\ \vdots & & \vdots & \vdots \\ b & \cdots & 1 & 0 \\ 1 & \cdots & 1 & 1 \end{pmatrix}, \tag{3.1.35}$$

which is necessarily non-singular. So V_b^{-1} exists and the $\boldsymbol{\mu}_j$ can be obtained from

$$(\boldsymbol{\mu}_b \ \cdots \ \boldsymbol{\mu}_0) = V_b^{-1}. \tag{3.1.36}$$

From (3.1.33) and (3.1.34), noting (3.1.29), $(Y(b) \ \cdots \ Y(0))(\boldsymbol{\mu}_b \cdots \boldsymbol{\mu}_0)$ is null. So, due to (3.1.36),

$$Y(t) = 0, \quad t = 0, \dots, b. \tag{3.1.37}$$

Next, we may write (3.1.3) as

$$Y_b(h+b) = Z(h+b) + (Z(b) \cdots Z(0))(\pi_b^h(h) \dots \pi_b^h(h+b))'; \tag{3.1.38}$$

and, putting $t = h + b$, (3.1.32) becomes [on using (3.1.1), (3.1.34) and (3.1.29)]

$$Y(h+b) = Z(h+b) - \sum_{j=0}^{b} (h+b)^j (Z(b) \cdots Z(0)) \boldsymbol{\mu}_j. \tag{3.1.39}$$

We are now going to show that $Y(h+b) = Y_b(h+b)$ or, equivalently,

$$-\sum_{j=0}^{b} (h+b)^j \boldsymbol{\mu}_j = (\pi_b^h(h) \dots \pi_b^h(h+b))', \quad h = 1, 2, \dots. \tag{3.1.40}$$

For that we need the following lemmas.

Lemma 3.1.2 *For $h \geq 1, \ b \geq 0$,*

$$\{\Delta^b(\frac{s^j}{h+s})\}_{s=0} = (-1)^{b-j} \frac{b! h^j}{(h+b)^{(b+1)}}, \ j = 0, \dots, b. \tag{3.1.41}$$

Proof First notice that, for any integer $j > 0$,

$$\frac{s^j}{h+s} = s^{j-1} \frac{h+s-h}{h+s} = s^{j-1} - \frac{hs^{j-1}}{h+s}.$$

Then, since Δ^b annihilates any polynomial in s of degree less than b, we have for $j \in \{0, \dots, b\}$:

$$\Delta^b(\frac{s^j}{h+s}) = -h\Delta^b(\frac{s^{j-1}}{h+s}) = \dots = (-h)^j \Delta^b(\frac{1}{h+s}). \tag{3.1.42}$$

Now, in view of (1.2.2) and (1.5.12),

$$(-1)^b\{\Delta^b(\frac{1}{h+s})\}_{s=0} = \sum_{i=0}^{b}(-1)^i \binom{b}{i} \frac{1}{h+i} = \frac{R(1)b!}{(h+b)^{(b+1)}}; \tag{3.1.43}$$

but (1.5.3) gives $R(1) = 1$, so (3.1.41) follows from (3.1.42) and (3.1.43). □

Lemma 3.1.3 *For $h \geq 1$, $b \geq 0$,*

$$(h+b)^j = \frac{(h+b)^{(b+1)}}{b!}\{\sum_{i=0}^{b}(-1)^i \binom{b}{i} \frac{(b-i)^j}{h+i}\},\ j = 0,\ldots,b. \tag{3.1.44}$$

Proof By first expanding $(b-s)^j$ and then using (3.1.41), we get

$$[\Delta^b\{\frac{(b-s)^j}{h+s}\}]_{s=0} = \sum_{i=0}^{j}\binom{j}{i} b^{j-i}[\Delta^b\{\frac{(-s)^i}{h+s}B\}]_{s=0}$$

$$= \sum_{i=0}^{j}\binom{j}{i} b^{j-i}(-1)^b \frac{b!h^i}{(h+b)^{(b+1)}} = (-1)^b \frac{b!(h+b)^j}{(h+b)^{(b+1)}}.$$

Next, use (1.2.2) to alternatively expand the $\Delta^b\{(b-s)^j/(h+s)\}$, when (3.1.44) results. □

Theorem 3.1.3 *Suppose $d \in \mathcal{N}$, $\{Z(t)\} \in \dot{I}_d$, $b \in \mathcal{N}$ and $b \geq d-1$; then the primary representation of $Z(t)$ of order b, $Y_b(t)$, $t = b+1, b+2, \ldots$, given by (3.1.3), has an alternative form $Y(t)$, given by (3.1.32), and hence the definition of $Y_b(t)$ can be extended to $t \in \{0,\ldots,b\}$ for which $Y_b(t) = 0$.*

Proof As shown above, we only need to prove (3.1.40).

In view of (1.5.15c), (3.1.44) can be alternatively written as

$$-(h+b)^j = (b^j \cdots 0^j)(\pi_b^h(h) \cdots \pi_b^h(h+b))'. \tag{3.1.45}$$

From (3.1.36), then (3.1.45) and (3.1.35),

$$(\mu_b \cdots \mu_0)(-(h+b)^b \cdots -(h+b)^0)'$$

$$= V_b^{-1}V_b(\pi_b^h(h) \cdots \pi_b^h(h+b))' = (\pi_b^h(h) \cdots \pi_b^h(h+b))', \tag{3.1.46}$$

that is (3.1.40).

Notice that $Y_b(t)$ was only defined in (3.1.3) for $t > b$. As $Y(t)$ and $Y_b(t)$ coincide when $t > b$, we may instead define $Y_b(t)$ by (3.1.32) to extend its range to $t \in \mathcal{N}$.

(3.1.37) then shows that, under this extension, $Y_b(t) = 0$ for $t \in \{0, \ldots, b\}$ [which is what we would get if we extended the range of (1.3.3) by these t]. □

When $b = d-1$, the summation in the RHS of (3.1.32) is $Z(t) - Y(t) = R(t)$, the initial component of $Z(t)$, from (3.1.14). So we have [cf (3.1.39)]:

Corollary 3.1.1 *The initial component $R(t)$ in (3.1.14) can be expressed as*

$$R(t) = \sum_{j=0}^{d-1} t^j (Z(d-1) \cdots Z(0)) \boldsymbol{\mu}_j, \quad t = 0, 1, \ldots, \tag{3.1.47}$$

where the $\boldsymbol{\mu}_j$, $j = 0, \ldots, d-1$, are given by (3.1.36). Both $P(t)$ and $Q(t)$ $[P(t) + Q(t) = Y_{d-1,d}(t)]$ are zero at $t = 0, \ldots, d-1$ [see (3.1.15) and (3.1.16)].

From (3.1.30), (3.1.34) and then (3.1.16),

$$\begin{aligned}
\pi_{\{h+b\}} &= \delta_{\{h+b\}} - \sum_{j=0}^{b} (h+b)^j (\delta_{\{b\}} \cdots \delta_{\{0\}}) \boldsymbol{\mu}_j \\
&= \delta_{\{h+b\}} + (\delta_{\{b\}} \cdots \delta_{\{0\}}) \sum_{j=0}^{b} \{-(h+b)^j \boldsymbol{\mu}_j\} \\
&= \delta_{\{h+b\}} + (\delta_{\{b\}} \ldots \delta_{\{0\}})(\pi_b^h(h) \cdots \pi_b^h(h+b))',
\end{aligned}$$

which gives:

Corollary 3.1.2 *For $h = t - b > 0$, the primary increment vector of order b has a decomposition corresponding to the measure defined by (3.1.30), where the λ_j are given by (3.1.34).*

3.2 Some Applications of the Representation and the Decomposition

3.2.1 An analysis of divergence rates

Suppose $\{Z(t)\} \in \dot{I}_d$, $d \in \mathcal{Z}^+$. Then, when $\{Z(t)\}$ is represented at the right order $b = d-1$, we get the decomposition (3.1.11). Now, we always assume the second-order moments of the $Z(t)$ exist for fixed finite t. So, almost all realizations of the stochastic series will have finite initial values, $Z(0)$, …, $Z(b)$; and, from (3.1.17),

$$R(t) = O(t^{d-1}) \quad a.s., \qquad \mathrm{Var}\{R(t)\} = O(t^{2d-2}). \tag{3.2.1}$$

Here, and in the sequel, "a.s." denotes "almost surely".

Next, denote the spectrum of $\{W(t)\}$ by $S^*(\omega') = \int_{-\pi}^{\omega'} s^*(d\omega)$, where

$$s^*(d\omega) = s(\omega)d\omega + \sum_r s_r \delta_{\omega-\omega_r}. \tag{3.2.2}$$

Here, $s(\omega)$ is the spectral density function of $\{U(t)\}$, the s_r are the spectral jumps of $\{V(t)\}$ and δ_ω is the Kronecker delta function. Then $\int_{-\pi}^{\pi} s^*(d\omega) = \gamma(0) = \text{Var}\{W(t)\} < \infty$ implies [see (3.1.13)]

$$\sum_r E\{|\xi_r|^2\} = \sum_r s_r < \infty. \tag{3.2.3}$$

Also, suppose that there is a $\dot{\omega} > 0$, such that

$$|\omega_r| > \dot{\omega} \qquad \forall\, \omega_r \in [-\pi, \pi] \quad (r \neq 0). \tag{3.2.4}$$

Then the $|1 - e^{i\omega_r}|$ are bounded away from zero and, in view of (3.1.10) and (3.2.3) (notice that $s_0 = 0$),

$$E\{|\sum_{r=-\infty}^{\infty} \frac{\xi_r e^{i\omega_r t}}{(1-e^{-i\omega_r})^d}|^2\} \le \sum_{r=-\infty}^{\infty} \frac{s_r}{|1-e^{i\omega_r}|^{2d}} < \infty$$

and

$$E\{|\sum_{r=-\infty}^{\infty} \frac{\xi_r e^{i\omega_r(d-1)}}{(1-e^{-i\omega_r})^{d-j}}|^2\} \le \sum_{r=-\infty}^{\infty} \frac{s_r}{|1-e^{-i\omega_r}|^{2d-2j}} < \infty.$$

So, from (3.1.25), we have

$$Q(t) = O(t^{d-1}) \quad a.s., \qquad \text{Var}\{Q(t)\} = O(t^{2d-2}). \tag{3.2.5}$$

In fact, $P(t)$ is the primary representation of $\check{Z}(t)$ of order $d-1$, where $\nabla^d \check{Z}(t) = U(t)$. From Theorem 2.4.1 (put $t = h + b = h + d - 1$), as $t \to \infty$, we have

$$\text{Var}\{P(t)\} = \frac{2\pi s(0)}{(2d-1)\{(d-1)!\}^2} t^{2d-1} + O(t^{2d-2+\eta}), \quad 0 \le \eta < 1. \tag{3.2.6}$$

Comparing (3.2.6) with the variances given in (3.2.1) and (3.2.5), we see that, when $s(0) > 0$, neither of the divergence rates of the variances of $R(t)$ and $Q(t)$ can attain the divergence rate of the variance of $P(t)$. That is, the regular component dominates the divergence behaviour of the whole series. Result (3.3.8), below, shows $\{P(t)\} = O(t^{d-1/2})$ *a.s.*, and this is a higher divergence rate than those of $R(t)$ and $Q(t)$ [which are both $O(t^{d-1})$ *a.s.*, from (3.2.1) and (3.2.5)]. So we may conclude that, for an $\tilde{I}_d$-series, the main source of the divergence behaviour is not some hidden trend, in general. This conclusion may appear counter-intuitive, but remember that our analysis assumes $E\{W(t)\} = 0$, so $\{Z(t)\}$ has no trend of degree exceeding $d-1$.

3.2.2 Overdifferenced models

In fact, for $\{Z(t)\} \in \tilde{I}_d$, $d \in \mathcal{Z}^+$, and any $b \in \mathcal{N}$, $b \geq d-1$, referring to (3.1.3), we can always write

$$Z(t) = Y_{bd}(t) + R_b(t) \quad (t = h + b,\ h = 1, 2, \ldots), \tag{3.2.7}$$

where $Y_{bd}(t)$ is the specific expression for the primary representation $Y_b(t)$ of $Z(t)$ of order b given by (2.2.5) [which is equivalent to (3.1.4)], and

$$R_b(t) = \frac{1}{b!}\sum_{j=0}^{b}(-1)^{b-j}\binom{b}{j}\frac{t^{(b+1)}}{t-j}Z(j). \tag{3.2.8}$$

When $b = d-1$, $R_b(t)$ is the $R(t)$ of (3.1.17), and $Y_{bd}(t) = P(t) + Q(t)$ is given by (3.1.15) and (3.1.16).

There is no representation for $Z(t)$ of order lower than $d-1$. While, from (2.2.3), we see that representing $Z(t)$ at an order b higher than $d-1$ means overdifferencing $Z(t)$. Now let us see what happens if $Z(t)$ is represented and "decomposed" at a higher order, say, $b = d$. Then

$$Z(t) = Y_{dd}(t) + R_d(t). \tag{3.2.9}$$

From (2.2.4) and noticing $t = h + d$,

$$\begin{aligned}
Y_{dd}(t) &= \{\frac{1}{d!}\sum_{j=0}^{t-d-1}(j+d)^{(d)}B^j(1-B)\}W(t) \\
&= \frac{1}{d!}\{\sum_{j=0}^{t-d}(j+d)^{(d)}B^j - t^{(d)}B^h - \sum_{j=-1}^{t-d-1}(j+d)^{(d)}B^{j+1}\}W(t) \\
&= \frac{1}{d!}[\sum_{j=0}^{t-d}\{(j+d)^{(d)} - (j+d-1)^{(d)}\}B^j - t^{(d)}B^h]W(t) \\
&= \{\frac{1}{(d-1)!}\sum_{j=0}^{t-d}(j+d-1)^{(d-1)}B^j - \frac{t^{(d)}}{d!}B^h\}W(t).
\end{aligned}$$

That is

$$Y_{dd}(t) = Y_{d-1,d}(t) - X_d(t), \tag{3.2.10}$$

where

$$X_d(t) = \{t^{(d)}/d!\}W(d). \tag{3.2.11}$$

On the other hand, from (3.2.8),

$$\begin{aligned} R_d(t) &= \frac{t^{(d)}}{d!}[\{\sum_{j=0}^{d}(-1)^{d-j}\binom{d}{j}(\frac{t-d}{t-j}-1)Z(j)\} + \nabla^d Z(d)] \\ &= \frac{t^{(d)}}{(d-1)!}\sum_{j=0}^{d-1}(-1)^{d-1-j}\binom{d-1}{j}\frac{1}{t-j}Z(j) + \frac{t^{(d)}}{d!}W(d). \end{aligned}$$

Thus

$$R_d(t) = R_{d-1}(t) + X_d(t). \tag{3.2.12}$$

From (3.2.10) and (3.2.12), we see that (3.2.9) can be written as

$$Z(t) = \{Y_{d-1,d}(t) - X_d(t)\} + \{R_{d-1}(t) + X_d(t)\}. \tag{3.2.13}$$

However, (3.2.11) gives

$$\mathrm{Var}\{X_d(t)\} = \{t^{(d)}/d!\}^2\gamma(0) = \{t^{2d}/(d!)^2\}\gamma(0) + O(t^{2d-1}), \tag{3.2.14}$$

which has a higher divergence rate than either $Y_{d-1,d}(t) = P(t) + Q(t)$ [see (3.2.5) and (3.2.6)] or $R_{d-1}(t)$ [see (3.2.1)]. So, decomposing $Z(t)$ at an order d, one greater than is necessary, involves introducing a term, $-X_d(t)$, into the sum of the regular and the deterministic components (the representation of just necessary order, $d-1$), and a term, $X_d(t)$, into the corresponding initial component. This results in the divergence rates of $Y_{dd}(t)$ and $R_d(t)$ being swollen by dominating terms which cancel on summing; and the same type of argument applies to all situations with higher degrees of overdifferencing (see the next subsection). Thus an overdifferenced "decomposition" is detrimental. However, the existence of such a term swelling the representation of any order $b \geq d$ becomes a touchstone for investigating whether a model of order b is overdifferenced or not. This provides insight as to why the polyvariogram, $\gamma_b(h) = E\{Y_b^2(h+b)\}/h^{2b}$, increases roughly linearly when $b = d-1$, and levels out when $b \geq d$.

3.2.3 Asymptotes of polyvariograms ($d < b+1$)

We are now able to prove Theorem 2.1.3. For $d \in \mathcal{Z}^+$, $\{Z(t)\} \in \tilde{I}_d$, $b \in \mathcal{N}$ and $b \geq d$, in exactly the same way that (3.2.10) was derived from (2.2.4), we get

$$Y_{bd}(h+b) = Y_{b-1,d}(h+b) - \frac{(h+b)^{(b)}}{b!}\nabla^b Z(b)$$

$$= Y_{b-2,d}(h+b) - \sum_{r=b-1}^{b} \frac{(h+b)^{(r)}}{r!}\nabla^r Z(r),$$

and, continuing recursively, this gives (for $d > 0$):

$$Y_{bd}(h+b) = Y_{d-1,d}(h+b) - \sum_{r=d}^{b} \frac{(h+b)^{(r)}}{r!}\nabla^r Z(r), \tag{3.2.15}$$

where [cf (2.2.5)]

$$Y_{d-1,d}(h+b) = \frac{1}{(d-1)!}\sum_{j=0}^{h+b-d}(j+d-1)^{(d-1)}W(h+b-j). \tag{3.2.16}$$

In the case of $d = 0$, the same technique leads to

$$Y_{b0}(h+b) = Y_{00}(h+b) - \sum_{r=0}^{b} \frac{(h+b)^{(r)}}{r!}\nabla^r Z(r). \tag{3.2.17}$$

Observe, from (3.2.15) and (3.2.17), that $r \geq d$ and put $\nabla^r Z(r) = \nabla^{r-d}\nabla^d Z(r) = \nabla^{r-d}W(r)$. Then it is easy to show that

$$\mathrm{Var}\{\nabla^r Z(r)\} = \{\Delta^{r-d}\nabla^{r-d}\gamma(h)\}_{h=0}, \tag{3.2.18}$$

$$\mathrm{Cov}\{\nabla^r Z(r), \nabla^s Z(s)\} = c_{rs} \quad \text{(a constant)}. \tag{3.2.19}$$

When $d > 0$, (2.2.24), (2.4.4) and condition (2.4.8) give

$$\mathrm{Var}\{Y_{d-1,d}(h+b)\} = v_{d-1,d}(h+b-d+1) = O(h^{2d-1}). \tag{3.2.20}$$

Next, from (3.2.16)

$$\mathrm{Cov}\{Y_{d-1,d}(h+b),\ \nabla^r Z(r)\} = \mathrm{Cov}\{Y_{d-1,d}(h+b),\ \nabla^{r-d}W(r)\}$$

$$= \frac{1}{(d-1)!}\sum_{j=0}^{h+b-d}(j+d-1)^{(d-1)}\nabla_r^{r-d}\gamma(h+b-j-r),$$

where ∇_r^{r-d} indicates the operator ∇^{r-d} operating on r; and, due to condition (2.4.8), the absolute value of this last expression is dominated by $ch^{d-1}, c > 0$. So,

$$\mathrm{Cov}\{Y_{d-1,d}(h+b),\ \nabla^r Z(r)\} = O(h^{d-1}). \tag{3.2.21}$$

Then, from (3.2.18),

$$\mathrm{Var}\{\frac{(h+b)^{(r)}}{r!}\nabla^r Z(r)\} = h^{2r}\{\Delta^{r-d}\nabla^{r-d}\gamma(h)\}_{h=0}/(r!)^2 + O(h^{2r-1}) \tag{3.2.22}$$

which can not exceed $O(h^{2b-2})$ when $r < b$. Finally, from (3.2.19),

$$\mathrm{Cov}\{\frac{(h+b)^{(r)}}{r!}\nabla^r Z(r),\ \frac{(h+b)^{(s)}}{s!}\nabla^s Z(s)\} = O(h^{r+s}) \tag{3.2.23}$$

which will not exceed $O(h^{2b-1})$ for $r \neq s$ and $r \leq b$, $s \leq b$. So (3.2.15) gives (2.4.21), on using (3.2.20) to (3.2.23).

When $d = 0$, the $v_{b0}(h)$ are the variances of the $Y_{b0}(h+b)$ given by (3.2.17). If $b > 0$, since $Y_{00}(h+b) = -W(0) + W(h+b)$ [see (2.2.31)], then (2.4.21) follows in the same way as for $d > 0$. If $b = 0$, $v_{00}(h) = \mathrm{Var}\{Y_{00}(h)\}$ is given in (2.2.32), which satisfies (2.4.21), since (2.4.8) implies that $\gamma(h) = o(h^{-1})$. The first part of Theorem 2.4.3 is thus proved.

We now turn to the proof of (2.4.22). For simplicity of notation, denote $\{\Delta^{b-d}\nabla^{b-d}\gamma(h)\}_{h=0}$ by a_{b-d}. For $b > 0$, in view of (1.4.13) and (1.4.17), and then using (2.4.21),

$$\begin{aligned}
\tilde{\kappa}_{bd}(h)/h &= b!(-1)^{b+1}\Sigma^b v_{bd}(h)/\{2(h+b)^{(b+1)}\} \\
&= (-1)^{b+1}a_{b-d}\Sigma^b[h^{(b-1)}/\{2(b!)\} + O(h^{b-2})] \\
&= (-1)^{b+1}a_{b-d}\Sigma^{b-1}[h^{(b)}/\{2b(b!)\} + O(h^{b-1})] \\
&\vdots \\
&= (-1)^{b+1}a_{b-d}h^{(2b-1)}/\{2b(b+1)\cdots(2b-1)(b!)\} + O(h^{2b-2}) \\
&= (-1)^{b+1}a_{b-d}h^{2b-1}/(2b)! + O(h^{2b-2}),
\end{aligned}$$

which gives (2.4.22) immediately. For $b = 0$, $\tilde{\kappa}_{00}(h)$ is given in (2.3.39), and again satisfies (2.4.22). That completes the proof of Theorem 2.4.3.

3.2.4 Forecasting

We assume that the degree of differencing, d, has been determined correctly. Then the decomposition theorems (Theorem 3.1.1 and Theorem 3.1.2) provide convenient tools for transforming nonstationary forecasting problems to stationary ones. In (3.1.14), $Q(t)$ and $R(t)$ are completely predictable. $R(t)$ is the simplest component, and its predicted values can be obtained directly from data, $Z(0), \ldots, Z(d-1)$, using (3.1.17). However, for $Q(t)$, we are in difficulty, since we usually do not know the ω_r and ξ_r. But, as $Q(t)$ is brought into $Z(t)$ from the sinusoids hidden in the hub series, and because we may assume in practice that there are only a finite number of ω_r, then ω_r and ξ_r can be estimated from observations on $\{W(t)\}$ [i.e., having $Z(0), \ldots, Z(n)$, the $W(t) = \nabla^d Z(t),\ t = d, \ldots, n$, can be obtained]. Notice that, although the ξ_r are random variables, they are fixed quantities in a given realization.

The classical theory for detecting and estimating the ω_r for a stationary series has a long history, and a good introduction to this is given by Priestley (1980). Recent development (Chen 1988a, b) shows that there are procedures for detecting all ω_r if the length of the observed series, n, is large enough, and that the ω_r can then be estimated with a precision of $O\{(n^{-3}\log\log n)^{1/2}\}$ almost surely. Once the ω_r have been estimated, the ξ_r can be easily estimated by least squares with its usual precision of $O_p(n^{-1/2})$.

In the literature, most authors are consciously aware of the existence of polynomial trends in I_d-series (but not of sinusoidal trends). The problem, however, is to estimate the trend in such a series. For instance, how does a regression procedure fare, when the divergence rate of the "error" $[P(t)]$ dominates the divergence rate of the "trend" $[Q(t) + R(t)]$? That is very hard to gauge. Having these decomposition theorems, when the ω_r and ξ_r are estimated from observations of the hub series (obtained from d-times differencing the data), (3.1.17) and (3.1.25) become explicit formulae for estimating trends in I_d-series.

In most commonly used models (such as ARIMA models), the absence of $Q(t)$ is assumed. [Should we allow the possibility of a non-null $Q(t)$, the decomposition theorems may be used to estimate $Q(t)$, which can then be subtracted from $Z(t)$.]

When there is no $Q(t)$, the problem of forecasting $Z(t)$ is reduced to that of forecasting $P(t)$, and we have $U(t) = W(t) = \nabla^d Z(t)$. It is more convenient to write (3.1.15) as

$$P(t) = \frac{1}{(d-1)!}\sum_{j=d}^{t}(d-1+t-j)^{(d-1)}U(j). \tag{3.2.24}$$

Let $t \geq l > t_0$ and $\hat{U}(l \mid t_0, d)$ denote the conditional expectation of $U(l)$ given $\{U(j),\ d \leq j \leq t_0\}$ [or, alternatively, the projection of $U(l)$ on the linear space generated by $U(j)$, $d \leq j \leq t_0$]. Then the best (or the best linear) forecast of $P(t)$, given $Z(0), \ldots, Z(t_0)$ [so that the $U(j) = \nabla^d Z(j)$, $j = d, \ldots, t_0$, are all known] is

$$\hat{P}(t \mid t_0, d) = \frac{1}{(d-1)!}\sum_{j=d}^{t_0}(d-1+t-j)^{(d-1)}U(j)$$

$$+\frac{1}{(d-1)!}\sum_{l=t_0+1}^{t}(d-1+t-l)^{(d-1)}\hat{U}(l \mid t_0, d) \tag{3.2.25}$$

and the best (or the best linear) forecast of $Z(t)$ is

$$\hat{Z}(t \mid t_0, 0) = \hat{P}(t \mid t_0, d) + R(t). \tag{3.2.26}$$

Where $R(t)$ is given by (3.1.17) or (3.1.17).

The forecast $\hat{U}(l \mid t_0, d)$ of a stationary series could be either parametric or non-parametric.

Example When $d = 1$, from (3.2.25) and (3.1.17),

$$\hat{P}(t \mid t_0, 1) = \sum_{j=1}^{t_0}\nabla Z(j) + \sum_{l=t_0+1}^{t}\hat{U}(l \mid t_0, 1), \quad R(t) = Z(0). \tag{3.2.27}$$

Then, from (3.2.26),

$$\hat{Z}(t \mid t_0, 0) = Z(t_0) + \sum_{l=t_0+1}^{t}\hat{U}(l \mid t_0, 1). \tag{3.2.28}$$

Certainly, in this case, (3.2.28) is well-known nd can be easily derived without using (3.2.25). However, for $d > 1$, (3.2.25) is useful theoretically and also convenient for computation.

3.3 The Divergence Rate of Realizations

3.3.1 The divergence rate theorem

In subsecton 3.2.1, we mentioned that, for $\{Z(t)\} \in \tilde{I}_d$, $d \in \mathcal{Z}^+$, when (3.2.3) and (3.2.4) hold and the spectral density, $s(\omega)$, of $\{U(t)\}$ satisfies

$$s(0) > 0, \tag{3.3.1}$$

then the divergence rate of the whole series is determined by $P(t)$ in the decomposition (3.1.14), while $Q(t)$ and $R(t)$ only contribute trends of lower order [see (3.2.1) and (3.2.5)]. However, so far, we have only obtained the divergence rate for the variance of $P(t)$ [(3.2.6)]. The next theorem will describe the divergence behaviour of almost all realizations of $\{P(t)\}$ (or, equivalently, of $\{Z(t)\}$ itself). For this purpose, rather than (3.1.7) and (3.1.8), we need the stronger conditions:

$$U(t) = \sum_{j=0}^{\infty} \alpha_j A(t-j), \quad \sum_{j=0}^{\infty} | \alpha_j | < \infty; \tag{3.3.2}$$

$$A(t) \quad \text{are independent,} \quad E\{A(t)\} = 0, \; E\{A^2(t)\} = \sigma^2; \tag{3.3.3}$$

$$\sup_t E\{| A(t) |^r\} < \infty, \quad \text{for some} \quad r > 2. \tag{3.3.4}$$

Theorem 3.3.1 *Suppose $\{Z(t)\} \in \tilde{I}_d$, $d \in \mathcal{Z}^+$, and the hub series $\{W(t)\}$ has the Wold decomposition (3.1.6). Then, if (3.2.4) holds [for $\{V(t)\}$] and (3.3.1) through (3.3.4) hold [for $\{U(t)\}$],*

$$\limsup_{t\to\infty} \frac{| Z(t) | (d-1/2)^{1/2}(d-1)!}{\{2\pi s(0) t^{2d-1} \log\log t\}^{1/2}} = 1 \quad a.s.\,. \tag{3.3.5}$$

Remark A stochastic series, $\{Z(t)\}$, should be written as $\{Z(t,\omega)\}$, a function defined on a product space $\mathcal{N} \times \Omega$. For a fixed $t \in \mathcal{N}$, $\{Z(t,\omega)\}$ is a random variable defined on a basic probability space, say $(\Omega, \mathcal{F}, P)$. Then, for a fixed $\omega \in \Omega$, $\{Z(t,\omega)\}$ is a realization. Let $t \to \infty$, and use the abbreviated notation "$f(Z(t)) \to 1$ a.s." to mean "$f(Z(t,\omega)) \to 1$ for all $\omega \in \Omega'$, where $\Omega' \subset \Omega$ and $P(\Omega') = 1$". □

(3.3.5) is a law of the iterated logarithm (LIL). Roughly, it tells us that, as $t \to \infty$,

$$Z(t) = O(t^{2d-1} \log\log t)^{1/2} \quad a.s.\,. \tag{3.3.6}$$

But (3.3.5) is much stronger than (3.3.6), as it gives the actual divergence rate (rather than just its order) which must be reached by almost all realizations.

In fact, Lai an Wei (1983) showed that (3.3.6) holds for the non-explosive autoregressive model,

$$Z(t) = \varphi_1 Z(t-1) + \ \ldots \ + \varphi_p Z(t-p) + A(t), \tag{3.3.7}$$

where $\varphi(\zeta) \equiv 1 - \varphi_1\zeta - \ldots - \varphi_p\zeta^p$ has none of its zeros within the unit circle and d is the highest multiplicity of any zero on the unit circle. Put $\varphi(\zeta) = u(\zeta)\phi(\zeta)$, where $u(\zeta)$ has all its roots on the unit circle and $\phi(\zeta)$ has all its roots outside the unit circle. Then we may write (3.3.7) as $u(B)Z(t) = \phi(B)^{-1}A(t)$. An $\tilde{I}_d$-series could be a special case of this, if $u(B) = \nabla^d$; but, then, the $\{\phi(B)^{-1}A(t)\}$ provides only a restricted case of a second-order stationary series, $\{W(t)\}$. That is, the set of the $\tilde{I}_d$-series and the set of non-explosive autoregressive series are distinct sets of stochastic series, which each include their non-empty intersection and other elements. So, Theorem 3.3.1 also shows that, for some non-explosive autoregressive models, the dominating order (3.3.6) is attainable; but I conjecture that this can only occur when $\varphi(\zeta)$ has its zeros of highest multiplicity on the unit circle at $\zeta = \pm 1$.

As $Z(t) = P(t) + Q(t) + R(t)$, in view of (3.2.1) and (3.2.5), proving (3.3.6) is equivalent to proving

$$\limsup_{t\to\infty} \frac{|\,P(t)\,|\,(d-1/2)^{1/2}(d-1)!}{\{2\pi s(0)t^{2d-1}\log\log t\}^{1/2}} = 1 \quad a.s. \tag{3.3.8}$$

3.3.2 Proof of the divergence rate theorem

The key idea of the proof is to put $P(t)$ [given by (3.2.21)] in the form

$$P(t) = \sum_{j=d}^{t} a_{tj} U(j), \quad t = d,\ d+1, \ldots, \tag{3.3.9}$$

$$a_{tj} = (d-1+t-j)^{(d-1)}/(d-1)!, \quad j = d, \ldots, t, \tag{3.3.10}$$

and then use the following theorem which is due to Chen (1990) (we modify its enunciation to suit the present context).

A Preliminary Theorem Suppose $U(t)$ satisfies (3.3.2), (3.3.3) and (3.3.4), and the spectral density function is denoted by $s(\omega)$. Assume $a_{tj} = 0$ if $j > t$; and, as $t \to \infty$,

$$a_t = \sum_{j=d}^{t} a_{tj}^2 \to \infty, \tag{3.3.11}$$

$$\sup_{j \geq d} a_{tj}^2 = o\{a_t(\log a_t)^{-\rho}\} \quad \text{for all } \rho > 0. \tag{3.3.12}$$

Assume also that there exist constants c_j and $q > 2/r$ such that, for $t > s \geq s_0 \geq d$,

$$\sum_{j=d}^{t}(a_{tj} - a_{sj})^2 \leq (\sum_{j=s+1}^{t} c_j)^q; \tag{3.3.13}$$

and, as $t \to \infty$,

$$(\sum_{j=s_0}^{t} c_j)^q = O(a_t). \tag{3.3.14}$$

Then,

$$\limsup_{t\to\infty} \mid P(t) \mid /(2\sigma^2 \tilde{a}_t \log\log \tilde{a}_t)^{1/2} \leq 1 \quad a.s., \tag{3.3.15}$$

where

$$\tilde{a}_t = \sigma^{-2} \int_{-\pi}^{\pi} s(\omega) \mid \sum_{j=d}^{t} a_{tj} e^{-i\omega j} \mid^2 d\omega. \tag{3.3.16}$$

Moreover: if, for every ζ $(0 < \zeta < \zeta_0)$, there exists a sequence of integers t_k (which may depend on ζ), $d < t_1 < t_2 < \ldots$, such that

$$\liminf_{k\to\infty} (t_k - t_{k-1})/t_k = \lambda > 0, \tag{3.3.17}$$

$$\limsup_{k\to\infty} \sum_{j=d}^{t_{k-1}} a_{t_k j}^2 / a_{t_k} < \zeta, \tag{3.3.18}$$

$$\limsup_{k\to\infty} \log\log a_{t_k} / \log k < 1 + \zeta, \tag{3.3.19}$$

$$\liminf_{k\to\infty} \log\log a_{t_k} / \log k > 0, \tag{3.3.20}$$

then the equality in (3.3.15) holds. □

We are now going to check these conditions for the a_{tj} given by (3.3.10), noting that $a_{tj} = 0$ when $j > t$. First, (3.3.11) and (3.3.12). In view of (2.2.26) and (2.4.5),

$$a_t = \frac{1}{\{(d-1)!\}^2} \sum_{j=d}^{t} \{(d-1+t-j)^{(d-1)}\}^2 = \frac{1}{\{(d-1)!\}^2} \sum_{j=0}^{t-d} \{(d-1+j)^{(d-1)}\}^2$$

$$= g_{d-1,d}(t-d+1,0) = t^{2d-1}/[(2d-1)\{(d-1)!\}^2] + O(t^{2d-2+\eta}), \tag{3.3.21}$$

where $0 \leq \eta < 1$. Obviously, (3.3.11) and (3.3.12) hold.

Next, consider (3.3.13) and (3.3.14). We can write

$$\sum_{j=d}^{t}(a_{t_j} - a_{s_j})^2 = \sum_{j=d}^{s}(a_{t_j} - a_{s_j})^2 + \sum_{j=s+1}^{t} a_{t_j}^2 . \tag{3.3.22}$$

Now, in the same way that (3.3.21) was derived, but using (2.4.4) rather than (2.4.5):

$$\sum_{j=s+1}^{t} a_{t_j}^2 = g_{d-1,d}(t-s,0) \leq \frac{(2d-2+t-s)^{2d-1}}{\{(d-1)!\}^2}.$$

Then it is easy to show that there is a constant, $c > 0$, such that

$$\sum_{j=s+1}^{t} a_{t_j}^2 \leq \frac{c}{2d-1}(t^{2d-1} - s^{2d-1}) = c\int_s^t x^{2d-2}dx \leq c\sum_{j=s+1}^{t} j^{2d-2}. \tag{3.3.23}$$

Write $m = d-1$ and $n = m+t-s$ $(> m)$. Then, due to (3.3.10),

$$\{(d-1)!\}^2\sum_{j=d}^{s}(a_{t_j} - a_{s_j})^2 = \sum_{j=0}^{s-d}\{(n+j)^{(d-1)} - (m+j)^{(d-1)}\}^2$$

$$= \sum_{i=0}^{s-d}\{(n+j)^{(d-1)}(n+j)^{(d-1)} - (n+j)^{(d-1)}(m+j)^{(d-1)}$$

$$-(m+j)^{(d-1)}(n+j)^{(d-1)} + (m+j)^{(d-1)}(m+j)^{(d-1)}\}; \tag{3.3.24}$$

and, using identity (2.4.1), we get

$$\{(d-1)!\}^2\sum_{j=d}^{s}(a_{t_j} - a_{s_j})^2$$

$$= \sum_{j=0}^{s-d}\Delta[\sum_{k=0}^{b}\frac{(-1)^k(d-1)^{(k)}}{(d+k)^{(k+1)}}\{(n+j)^{(d-1-k)}(n+k+j)^{(d+k)}$$

$$-(n+j)^{(d-1-k)}(m+k+j)^{(d+k)} - (m+j)^{(d-1-k)}(n+k+j)^{(d+k)}$$

$$+(m+j)^{(d-1-k)}(m+k+j)^{(d+k)}\}]$$

$$= \sum_{k=0}^{b}\frac{(-\ )^k(d-1)^{(k)}}{(a+k)^{(k+1)}}[\{t^{(d-1-k)}t^{(d+k)} - n^{(d-1-k)}n^{(d+k)}\} - t^{(d-1-k)}s^{(d+k)}$$

$$-\{s^{(d-1-k)}t^{(d+k)} - m^{(d-1-k)}n^{(d+k)}\} + s^{(d-1-k)}s^{(d+k)}]$$

$$< \sum_{k=0}^{b}\frac{(d-1)^{(k)}}{(d+k)^{(k+1)}}\{t^{(d-1-k)}t^{(d+k)} - s^{(d-1-k)}t^{(d+k)}\}$$

$$
\begin{aligned}
&< \sum_{k=0}^{b} \frac{(d-1)^{(k)}}{(d+k)^{(k+1)}}\{(t+d+k)^{(2d-1)} - s^{(2d-1)}\} \\
&= (2d-1)\sum_{k=0}^{b}\{\frac{(d-1)^{(k)}}{(d+k)^{(k+1)}}\sum_{j=s+1}^{t+d+k} j^{(2d-2)}\} \\
&\le \{(2d-1)\sum_{k=0}^{b}\frac{(d-1)^{(k)}}{(d+k)^{(k+1)}}\}\sum_{j=s+1}^{t+d+b} j^{2d-2} \\
&= c'\{(d-1)!\}^2 \sum_{j=s+1}^{t+d+b} j^{2d-2}.
\end{aligned} \tag{3.3.25}
$$

say. So, noticing that

$$
\sum_{j=t+1}^{t+d+b} j^{2d-2} < (d+b)(t+d+b)^{2d-2} \le (d+b)^{2d-1}t^{2d-2} \le c'' \sum_{j=s+1}^{t} j^{2d-2},
$$

(3.3.25) gives

$$
\sum_{j=d}^{s}(a_{tj} - a_{sj})^2 \le c'(1+c'') \sum_{j=s+1}^{t} j^{2d-2} = c''' \sum_{j=s+1}^{t} j^{2d-2}, \tag{3.3.26}
$$

say.

Then, writing

$$
c_j = (c + c''')j^{2d-2} \tag{3.3.27}
$$

and combining (3.3.22), (3.3.23) and (3.3.26), we see that (3.3.13) holds with $q = 1$; and, in view of (3.3.21), (3.3.14) also holds. That means (3.3.15) is true when the $P(t)$ are given by (3.3.9) with (3.3.10).

Now, from (3.3.10), (3.3.16) becomes

$$
\begin{aligned}
\tilde{a}_t &= \sigma^{-2}\int_{-\pi}^{\pi} s(\omega) \mid \sum_{j=d}^{t}(d-1+t-j)^{(d-1)}e^{-i\omega j}/(d-1)! \mid^2 d\omega \\
&= \sigma^{-2}\int_{-\pi}^{\pi} s(\omega) \mid \sum_{j=d}^{t}(d-1+t-j)^{(d-1)}e^{i\omega(t-j)}/(d-1)! \mid^2 d\omega \\
&= \sigma^{-2}\int_{-\pi}^{\pi} s(\omega) \mid \sum_{j=0}^{t-d}\frac{(d-1+j)^{(d-1)}}{(d-1)!}e^{i\omega j} \mid^2 d\omega.
\end{aligned}
$$

Comparing this with (2.2.5) (put $t-d = h-1$, $d-1 = b$) and noticing that

$$
\gamma(j) = \int_{-\pi}^{\pi} s(\omega)e^{i\omega j}d\omega, \tag{3.3.28}
$$

we get

$$\tilde{a}_t = \sigma^{-2}E\{Y_{d-1,d}^2(t)\} = \sigma^{-2}v_{d-1,d}(t-d+1), \tag{3.3.29}$$

on using (2.3.2). In view of (2.4.11),

$$\tilde{a}_t = \sigma^{-2}2\pi s(0)t^{2d-1}/[(2d-1)\{(d-1)!\}^2] + O(t^{2d-2+\eta}), \tag{3.3.30}$$

where $0 \le \eta < 1$. Replace $\tilde{a}_t$ in (3.3.15) by (3.3.30) and notice that, from (3.3.30), $\log\log \tilde{a}_t/\log\log t \to 1$. This gives (3.3.8), except that as yet we only have " ≤ 1", rather than " $= 1$".

For a fixed $\mu > 1$, let t_k be the integral part of μ^k. Then, from (3.3.10) and (2.2.26),

$$\begin{aligned}
\sum_{j=d}^{t_{k-1}} a_{t_k j}^2 &= \frac{1}{\{(d-1)!\}^2}\sum_{j=d}^{t_{k-1}}\{(d-1+t_k-j)^{(d-1)}\}^2 \\
&= \frac{1}{\{(d-1)!\}^2}\sum_{j=t_k-t_{k-1}}^{t_k-d}\{(d-1+j)^{(d-1)}\}^2 \\
&= \frac{1}{\{(d-1)!\}^2}\Big(\sum_{j=0}^{t_k-d} - \sum_{j=0}^{t_k-t_{k-1}-1}\Big)\{(d-1+j)^{(d-1)}\}^2 \\
&= g_{d-1,d}(t_k-d+1,0) - g_{d-1,d}(t_k-t_{k-1},0).
\end{aligned}$$

Thus, by (2.4.5) with $b = d-1$ and $t_k \simeq \mu^k$, we have

$$\sum_{j=d}^{t_{k-1}} a_{t_k j}^2 = \frac{\mu^{2dk-k} - \mu^{2dk-k}(1-\mu^{-1})^{2d-1} + O(\mu^{2dk-2k})}{(2d-1)\{(d-1)!\}^2}, \tag{3.3.31}$$

while (3.3.21) gives

$$a_{t_k} = \sum_{j=d}^{t_k} a_{t_k j}^2 = \frac{\mu^{2dk-k} + O(\mu^{2dk-2k+\eta k})}{(2d-1)\{(d-1)!\}^2}. \tag{3.3.32}$$

According to the definition of t_k, as $k \to \infty$,

$$(t_k - t_{k-1})/t_k \simeq (\mu^k - \mu^{k-1})/\mu^k \to 1-\mu^{-1} > 0;$$

so (3.3.17) holds. From (3.3.31) and (3.3.32), when $k \to \infty$,

$$\sum_{j=d}^{t_{k-1}} a_{t_k j}^2/a_{t_k} = 1-(1-\mu^{-1})^{2d-1} + o(1),$$

which can be made less than any $\zeta > 0$ by choosing μ sufficiently large; so (3.3.18) holds. Again from (3.3.32),

$$\log\log a_{t_k} = \log\{k \log \mu^{2d-1} + O(1)\} = \log k + O(1),$$

and (3.3.19) and (3.3.20) follow immediately. Thus, we have verified the four extra conditions of the preliminary theorem, which establishes the required equality in (3.3.8). $\square$

Chapter 4

Inference from Polyvariograms

4.1 Sample Polyvariograms and a Graphical Procedure

4.1.1 Definitions

Two types of polyvariogram were defined by (1.4.20) and (1.4.21), and we now consider their estimation. Given $b \in \mathcal{N}$, when $\{Z(t)\} \in \tilde{I}_d$, $d \in \{0, \ldots, b+1\}$, then the variogram of $\{Z(t)\}$ of order b is defined by (1.4.8). Also, for a fixed h, $\{Y_b(h+b,t)\}$ in (1.4.8) is stationary. So, if we have observations, $Z(0), \ldots, Z(n)$, $n > h+b$, we may first obtain $Y_b(h+b,t)$, $t = 0, \ldots, n-(h+b)$, and then use the *sample variogram*

$$\hat{v}_b(h) = \frac{1}{n-h-b+1} \sum_{t=0}^{n-h-b} Y_b^2(h+b,t), \quad h = 1, \ldots, n-b, \tag{4.1.1}$$

and the *sample polyvariogram* (SPV)

$$\hat{\gamma}_b(h) = \hat{v}_b(h)/h^{2b}, \quad h = 1, \ldots, n-b, \tag{4.1.2}$$

to estimate $v_b(h)$ and $\gamma_b(h)$, respectively.

In the problem of determining d, the correct degree of differencing, we will consider the *scaled polyvariogram*

$$\rho_b(h) = \gamma_b(h)/\gamma(0), \quad h = 1, 2, \ldots, \tag{4.1.3}$$

where $\gamma(0)$ is the variance of the hub series, $\{W(t)\}$, [see (1.1.5)]; and, correspondingly, we define the *scaled sample polyvariogram* (SSPV) as

$$\hat{\rho}_b(h) = \hat{\gamma}_b(h)/\hat{\gamma}(0), \quad h = 1, \ldots, n-b. \tag{4.1.4}$$

Here $\hat{\gamma}(0)$ is the sample variance of $\{W(t)\}$, the value at $j = 0$ of the sample autocovariance function (SACVF), $\hat{\gamma}(j)$, which we now define as

$$\hat{\gamma}(j) = \frac{1}{n-m} \sum_{t=m+1}^{n} W(t)W(t-j), \quad j = 0, 1, \ldots, m-d+1; \tag{4.1.5}$$

where m ($\geq d-1$) is an arbitrarily chosen fixed integer and we require $j \leq m-d+1$ since, given $Z(0), \ldots, Z(n)$, we can only obtain $W(d) \ldots, W(n)$. The typical definition of $\hat{\gamma}(j)$ in the literature is

$$\hat{\gamma}(j) = \frac{1}{n-d+1} \sum_{t=j+d}^{n} W(t)W(t-j), \quad j = 0, 1, \ldots, n-d; \tag{4.1.6}$$

but (4.1.5) is more convenient in our later discussion and, when j is restricted to the range stated there, its asymptotic properties are identical to those of (4.1.6).

From (1.4.13), (1.4.19) and (1.4.21), when $b \in \mathcal{Z}^+$ and $d \in \{0, \ldots, b+1\}$, the modified Cressie polyvariogram is

$$\gamma_b^*(h) = \frac{b}{2} \sum_{j=1}^{h-b} \frac{(h-1-j)^{(b-1)}}{(b+j)^{(b+1)}} v_b(j)/h^{2b-1}, \quad h = b+1, b+2, \ldots; \tag{4.1.7}$$

so $\gamma_b^*(\cdot)$ can be estimated by

$$\hat{\gamma}_b^*(h) = \frac{b}{2} \sum_{j=1}^{h-b} \frac{(h-1-j)^{(b-1)}}{(b+j)^{(b+1)}} \hat{v}_b(j)/h^{2b-1}, \quad h = b+1, \ldots, n, \tag{4.1.8}$$

which is the SPV for this type of polyvariogram. The *scaled polyvariogram* and the SSPV are this time defined by, respectively,

$$\rho_b^*(h) = \gamma_b^*(h)/\gamma(0), \quad h = b+1, b+2, \ldots, \tag{4.1.9}$$

and

$$\hat{\rho}_b^*(h) = \hat{\gamma}_b^*(h)/\hat{\gamma}(0), \quad h = b+1, \ldots, n. \tag{4.1.10}$$

When $b = 0$, since $\gamma_0^*(h) = \gamma_0(h)/2$, all the results concerning $\gamma_0^*(h)$ are clear from those for $\gamma_0(h)$.

Although $\hat{\gamma}_b(h)$ and $\hat{\gamma}_b^*(h)$ may be defined for h up to $n-b$ and n, respectively, we usually require $h \ll n$. In practice, for a graphical display, one may bring h near to n but, then, little may be confidently inferred from the SPV values at the later

h. When we make statistical inferences, the theoretical underpinning even requires $h \le k$ as $n \to \infty$, where k is a fixed number, chosen in advance.

As our other estimators can be expressed in terms of $\hat{v}_b(h)$ [see (4.1.2), (4.1.4), (4.1.8) and (4.1.10)], we consider $\hat{v}_b(h)$ to be the most fundamental estimator; and, in view of (1.4.6), (4.1.1) can be rewritten as

$$\hat{v}_b(h) = \frac{1}{n-h-b+1} \sum_{t=0}^{n-h-b} \{\mathbf{Z}'_{h+b}(t)\boldsymbol{\pi}_b^h\}^2, \quad h = 1, \ldots, n-b. \tag{4.1.11}$$

Cressie (1988) used an alternative estimator (when $b = 0, 1, 2$) which, with our notation, can be written as

$$\hat{v}_b(h) = \frac{1}{2(n-h-b+1)} \sum_{t=0}^{n-h-b} [\{\mathbf{Z}'_{h+b}(t)\boldsymbol{\pi}_b^h\}^2 + \{\mathbf{Z}'_{h+b}(t)\dot{\boldsymbol{\pi}}_b^h\}^2], \tag{4.1.12}$$

where $\dot{\boldsymbol{\pi}}_b^h$ is a vector obtained by reversing the order of all the elements of $\boldsymbol{\pi}_b^h$. (4.1.12) might have smaller variance than that of (4.1.11), so one might like to use (4.1.12) in a graphical investigation. But for statistical testing, a small reduction in variance is immaterial and so we prefer to use the simpler (4.1.11) in the development of our statistical inference theory.

4.1.2 Unbiasedness

Given $b \in \mathcal{N}$, all the above estimators only depend on the series data and b. For $\{Z(t)\} \in \tilde{I}_d$, $d \in \{0, \ldots, b+1\}$, we may expect distinct expressions corresponding to different d, so we introduce the specific d as a second suffix, as we did previously for their population quantities, giving $\hat{v}_{bd}(h)$, $\hat{\gamma}_{bd}(h)$ and so on. In view of (2.2.6), with $h+b$ replaced by $h+b+t$, and (4.1.1),

$$\hat{v}_{bd}(h) = \sum_{r=0}^{h+b-d} \sum_{s=0}^{h+b-d} \pi_{bd}^h(r)\pi_{bd}^h(s)\hat{\gamma}(h, r, s), \quad h = 1, \ldots, n-b, \tag{4.1.13}$$

where

$$\hat{\gamma}(h, r, s) = \frac{1}{n-h-b+1} \sum_{t=0}^{n-h-b} W(h+b+t-r)W(h+b+t-s). \tag{4.1.14}$$

Obviously, $E\{\hat{\gamma}(h, r, s)\} = \gamma(r-s)$; and then, in view of (2.2.24) and (2.2.25),

$$E[\hat{v}_{bd}(h)] = \sum_{r=0}^{h+b-d} \sum_{s=0}^{h+b-d} \pi_{bd}^h(r)\pi_{bd}^h(s)\gamma(r-s) = v_{bd}(h), \ h = 1, \ldots, n-b. \tag{4.1.15}$$

[The $\hat{v}_b(h)$ defined by (4.1.12) is also unbiased.] From (1.4.20), (4.1.2), (4.1.7), (4.1.8) and (4.1.15), we have

$$E\{\hat{\gamma}_{bd}(h)\} = \gamma_{bd}(h), \quad h = 1, \ldots, n-b, \tag{4.1.16}$$

$$E\{\hat{\gamma}^{*}_{bd}(h)\} = \gamma^{*}_{bd}(h), \quad h = b+1, \ldots, n. \tag{4.1.17}$$

So $\hat{v}_b(h)$, $\hat{\gamma}_b(h)$ and $\hat{\gamma}^{*}_b(h)$ are unbiased estimators of, respectively, $v_b(h)$, $\gamma_b(h)$ and $\gamma^{*}_b(h)$.

We now consider the SSPV. It is well-known (see, say, Fuller 1976, Theorem 6.2.3) that, for a stationary series, $\{W(t)\}$, under very mild conditions:

$$E\{\hat{\rho}(j)\} = \rho(j) + O(1/n), \quad 1 \leq j \leq m-d+1; \tag{4.1.18}$$

where m $(\geq d-1)$ is an arbitrarily chosen, but fixed, number and

$$\rho(j) = \gamma(j)/\gamma(0) \tag{4.1.19}$$

and

$$\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0) \tag{4.1.20}$$

are, respectively, the ACF and SACF of $\{W(t)\}$. Although, in the literature, (4.1.18) is proved for $\hat{\gamma}(j)$ defined by (4.1.6) (see, again, Fuller 1976, Theorems 6.2.1 and 6.2.3), it can be similarly proved when the $\hat{\gamma}(j)$ are defined by (4.1.5) or when the $\hat{\gamma}(j)$ are replaced by $\hat{\gamma}(h, r, s)$, defined by (4.1.14), with $|\, r - s \,| = j$. In this latter case, for all large n, from (4.1.13), followed by (4.1.18) with $m = k+b-1$, and finally (4.1.15), we have

$$E\{\hat{v}_{bd}(h)/\hat{\gamma}(0)\} = v_{bd}(h)/\gamma(0) + O(1/n), \quad 1 \leq h \leq k, \tag{4.1.21}$$

and hence [using (1.4.20), (4.1.2) to (4.1.4), and (4.1.7) to (4.1.10)]:

$$E\{\hat{\rho}_{bd}(h)\} = \rho_{bd}(h) + O(1/n), \quad 1 \leq h \leq k, \tag{4.1.22}$$

$$E\{\hat{\rho}^{*}_{bd}(h)\} = \rho^{*}_{bd}(h) + O(1/n), \quad b+1 \leq h \leq b+k \tag{4.1.23}$$

(for all large n). That is, the SSPVs are asymptotically unbiased when h is restricted to an appropriate fixed range of small consecutive integers; and the bias for finite-lengthed series is then $O(1/n)$.

4.1.3 A graphical procedure for determining the degree of differencing

The asymptotes of polyvariograms were given in Section 2.4. The key point is: if $\{Z(t)\} \in \tilde{I}_d$, then $\gamma_b(h)$ and $\gamma_b^*(h)$ have linear increasing trends, when $b = d-1$, but level out when $b \geq d$. This marked distinction, in theoretical behaviour, indicates the correct degree of differencing. Also, as we have just shown, the SPVs are unbiased estimators of the polyvariograms, so it is reasonable to expect that the SPVs will reflect the shapes of the theoretical polyvariograms. Here we propose the following Cressie-type procedure fo: identifying d.

Step 1 Guess an upper bound for d, d_0 say. In practice, we may choose $d_0 = 3$ or 4 in agreement with Box and Jenkins (1976) — who expect d to rarely exceed 2 for real data.

Step 2 For each of $b = d_0, d_0 - 1, \ldots$ (stopping at the $\hat{d} - 1$ suggested by step 3, below), use (4.1.11) or (4.1.12) and (4.1.2) to calculate $\hat{\gamma}_b(h)$ [or (4.1.8), to calculate $\hat{\gamma}_b^*(h)$] for $h = 1, \ldots, k$ [or $h = b+1, \ldots, b+k$]. The choice of k is debatable: it should be large enough to ensure that we can determine whether or not the polyvariogram levels out; but, for h approaching n, the SPV may start to fluctuate wildly. When n is large, it should be simple to choose a suitable k; but, for moderate n, some conservative upper limit on k, say $n/2$, may appear needed. However, as the computation is not costly, it would seem sensible (while one is gaining experience of the procedure) to go as far as is convenient and perhaps to the limit of $k = n - b$ [or $k = n$ for $\hat{\gamma}_b^*(h)$]; but bear in mind that the later behaviour of the SPV will probably need to be disregarded.

Step 3 If we find a particular b, $\hat{d} - 1$ say, such that, for $d_0 \geq b \geq \hat{d}$, $\hat{\gamma}_b(h)$ [or $\hat{\gamma}_b^*(h)$] levels out with increasing h, but, for $b = \hat{d} - 1$, $\hat{\gamma}_b(h)$ [or $\hat{\gamma}_b^*(h)$] appears to have a positive sloping asymptote, then $\hat{d}$ will be our estimate of d. If $\hat{\gamma}_0(h)$ [or $\hat{\gamma}_0^*(h)$] still levels out, then the estimate of d is zero. □

In fact, the basic idea of the above procedure was introduced by Cressie (1988). But, our suggestion of starting b from some perceived safe upper bound d_0 and then stopping the procedure as soon as an SPV demonstrates a positive sloping asymptote, which is contrary to the conventional routine of trying the b's in consecutive *increasing* order, can avoid the risk of returning aberrant $\hat{\gamma}_b(h)$ [or $\hat{\gamma}_b^*(h)$] for b that are too small (as the polyvariograms are undefined when $b < d-1$). For example, consider the low-order SSPVs shown in Figure 2 of Cressie (1988). These suggest that $d = 2$, as the scaled sample quadvariogram ($b = 2$) levels out but the scaled sample linvariogram ($b = 1$) has a positive sloping asymptote. But, having made this judgement, the implication is that the displayed scaled sample semivariogram ($b = 0$) does not reflect any theoretical scaled semivariogram (which would then be undefined), as indeed Cressie noted. [However, these comments do not preclude the possibility that some such $\hat{\gamma}_b(h)$ may, nonetheless, eventually prove to be useful extra tools for determining d. Compare using the SACFs for b-times differenced $\tilde{I}_d$-series, with $d \geq 1$.]

When we get $\hat{d}$, we can next verify the acceptability of this value by checking that the $\hat{\gamma}_{b\hat{d}}(h)$ [or $\hat{\gamma}^*_{b\hat{d}}(h)$] then (approximately) have the properties described by Corollary 2.4.3, when $b \geq \hat{d}$, and by Corollary 2.4.1 (or Corollary 2.4.2) when $b = \hat{d} - 1$. In the fomulae of these corollaries, $\gamma(j)$ can be estimated by $\hat{\gamma}(j)$ and $s(0)$ can be estimated via the Bartlett window:

$$\hat{s}(0) = \sum_{j=-j(n)}^{j(n)} \left(1 - \frac{|j|}{j(n)}\right)\hat{\gamma}(j), \tag{4.1.24}$$

where $j(n) = O(n^{1/3})$, or with other windows (see Priestley 1981). If $\hat{d}$ is a correct estimate of d and n is large enough, these corollaries should be roughly satisfied, when the unknown quantities in the formulae are replaced by their estimates.

4.2 Asymptotic Properties of Sample Polyvariograms

4.2.1 An alternative formula for sample variograms

We have already given a formula for sample variograms, (4.1.13), from which the formulae for the SPVs and SSPVs follow. Although the factor $\hat{\gamma}(h, r, s)$ in (4.1.13) is an unbiased estimator of $\gamma(r - s)$, it depends on three independent parameters which is not convenient for deriving asymptotic properties beyond unbiasedness. So we consider the difference between $\hat{\gamma}(h,r,s)$ and $\hat{\gamma}(r-s)$ [$\hat{\gamma}(i)$ is defined by (4.1.5)], which will lead to a more useful formula for $\hat{v}_{bd}(h)$, (4.2.6) below.

Lemma 4.2.1 *Suppose the $W(t)$ are identically distributed and $E\{|W(t)|^q\} < \infty$ for a constant $q > 0$; then, as $t \to \pm\infty$,*

$$W(t) = o(|t|^{1/q}) \cdot a.s.. \tag{4.2.1}$$

Proof Put $X(t) = |W(t)|^q$, so that the $X(t)$ are non-negative and identically distributed. Denote their distribution function by $F(x)$, giving $\int_0^\infty xF(dx) < \infty$. Then, for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{t=1}^{\infty} Pr\{|W(t)|/t^{1/q} > \varepsilon^{1/q}\} = \sum_{t=1}^{\infty} Pr\{X(t) > \varepsilon t\} \\
=\ & \sum_{t=1}^{\infty} \int_{\varepsilon t}^{\infty} F(dx) = \sum_{t=1}^{\infty}\sum_{k=t}^{\infty} \int_{\varepsilon k}^{\varepsilon(k+1)} F(dx) \\
\leq\ & \sum_{t=1}^{\infty}\sum_{k=t}^{\infty} \frac{1}{\varepsilon k} \int_{\varepsilon k}^{\varepsilon(k+1)} xF(dx) = \sum_{k=1}^{\infty} \frac{k}{\varepsilon k} \int_{\varepsilon k}^{\varepsilon(k+1)} xF(dx) \\
=\ & \frac{1}{\varepsilon} \int_{\varepsilon}^{\infty} xF(dx) < \infty.
\end{aligned}$$

Then (4.2.1) follows from the Borel-Cantelli lemma on letting $\varepsilon \to 0$. □

Lemma 4.2.2 *Suppose the $W(t)$ are strictly stationary with zero mean, $E\{W^4(t)\} < \infty$, and $\hat{\gamma}(i)$ and $\hat{\gamma}(h,r,s)$ are defined by (4.1.5) and (4.1.14), respectively, with $m \in \mathcal{N}$ fixed. Also, let $k \in \mathcal{Z}^+$ be fixed and $\mathcal{D}$ be the set of (h,r,s) satisfying*

$$1 \leq h \leq k, \quad |r-s| = i, \quad r,s \in \{0,\ldots,h+b-d\}, \quad 0 \leq i \leq m-d+1. \tag{4.2.2}$$

Then,

$$\max_{\mathcal{D}} |\hat{\gamma}(h,r,s) - \hat{\gamma}(i)| = o(n^{-1/2}) \quad a.s.. \tag{4.2.3}$$

Proof On using Lemma 4.2.1, for any fixed r' and s', since $E\{|W(t+r')W(t+s')|^2\} < \infty$, we have

$$W(t+r')\,W(t+s') = o(|t|^{1/2}) \quad a.s.. \tag{4.2.4}$$

Consider a fixed point (h,r,s) in $\mathcal{D}$; then, looking at (4.1.14),

$$\frac{1}{n-m} \sum_{t=0}^{n-h-b} W(h+b+t-r)W(h+b+t-s) - \hat{\gamma}(h,r,s) = o(n^{-1/2}) \quad a.s., \tag{4.2.5}$$

since all summands in the above summation are of order $o(n^{1/2})$ *a.s.*, but there are at most n summands and $1/(n-m) - 1/(n-h-b+1) = O(n^{-2})$. Next, for the following pair of summations,

$$\sum_{t=0}^{n-h-b} W(h+b+t-r)W(h+b-t-s) \quad \text{and} \quad \sum_{t=m+1}^{n} W(t)W(t-i),$$

there can only be a finite number of summands that do not appear in both. Thus, in view of (4.2.4), the difference between these two summations is $o(n^{1/2})$ *a.s.*. Finally, divide these summations by $n-m = O(n)$. Then, in view of (4.2.5), we get $\hat{\gamma}(h,r,s) - \hat{\gamma}(i) = o(n^{-1/2})$ *a.s.*. But $\mathcal{D}$ is a finite set, so (4.2.3) follows.

Theorem 4.2.1 *Suppose $\{Z(t)\} \in \tilde{I}_d$, with its hub series, $\{W(t)\}$, satisfying the conditions in Lemma 4.2.2. Futher, let $b \in \mathcal{N}$ and $b \geq \max\{d-1, 0\}$, and define $\hat{\gamma}(i)$ by (4.1.5) with, say, $m = k+b-1$. Then*

$$\hat{v}_{bd}(h) = \sum_{i=0}^{h+b-d} g_{bd}(h,i)\, \hat{\gamma}(i) + o(n^{-1/2}) \quad a.s., \quad 1 \leq h \leq k, \tag{4.2.6}$$

where $g_{bd}(h,i)$ is given by (2.2.25) and $o(n^{-1/2})$ is uniform in h.

Proof Due to (4.1.13) and (4.2.3), and $\pi^h_{bd}(r)$ being bounded in the range $0 \leq r \leq k+b-d$,

$$\hat{v}_{bd}(h) = \sum_{r=0}^{h+b-d} \sum_{s=0}^{h+b-d} \pi^h_{bd}(r)\pi^h_{bd}(s)\hat{\gamma}(r-s) + o(n^{-1/2}) \ a.s. \tag{4.2.7}$$

in the range $1 \leq h \leq k$. Then (4.2.6) follows from (2.2.25); and, from (4.2.3), we see that $o(n^{-1/2})$ is uniform in h. □

4.2.2 The asymptotic distributions of the SPVs and SSPVs

From (4.2.6), we see that, in the range of h, $1 \leq h \leq k$, the sample variogram can be explicitly expressed in terms of the sample autocovariance function of the hub series except for an $o(n^{-1/2})$ remainder which converges almost surely and uniformly in h. This enables us to use results for the SCVF, which have been established for stationary series, to derive the asymptotic properties of sample variograms and hence of SPVs and SSPVs. In this subsection, we consider the asymptotic distributions when the hub series, $\{W(t)\}$, is another form of general linear series:

$$W(t) = \sum_{j=-\infty}^{\infty} \alpha_j A(t-j), \quad \sum_{j=-\infty}^{\infty} |\alpha_j| < \infty, \tag{4.2.8}$$

where

$$A(t) \text{ are } i.i.d., \quad E\{A(t)\} = 0, \quad E\{A^2(t)\} = \sigma^2. \tag{4.2.9}$$

[Evidently, (4.2.9) is more restritive than (3.1.8); and it can be proved that (4.2.8) and (4.2.9) imply (3.1.7) and (3.1.8).] Throughout the rest of this thesis, $\{A(t)\}$ will always denote a white noise series, restricted by (4.2.9).

For a series, $\{W(t)\}$, given by (4.2.8) and (4.2.9), [referring to (4.1.5), (4.1.20) and (4.1.19)] write

$$\tilde{\gamma}(j) = \hat{\gamma}(j) - \gamma(j), \quad \tilde{\rho}(j) = \hat{\rho}(j) - \rho(j). \tag{4.2.10}$$

Then we have the following well-known theorem (for instance, see Anderson 1971, pp 478 & 489). Notice that, in this theorem, $\hat{\gamma}(j)$ may be defined by (4.1.5) or (4.1.6), or any other asymptotically equivalent form.

A Preliminary Theorem Let $\{W(t)\}$ be a series given by (4.2.8) and (4.2.9), and also assume $E\{A^4(t)\} < \infty$. Then, for a fixed $p \in \mathcal{N}$, as $n \to \infty$,

$$n^{1/2}(\tilde{\gamma}(0) \ \cdots \ \tilde{\gamma}(p))' \xrightarrow{L} N(\mathbf{0}, \Gamma_p), \tag{4.2.11}$$

where Γ_p is the $(p+1) \times (p+1)$ asymptotic variance-covariance matrix for the serial covariances, whose elements involve the 4th cumulant of $A(t)$.

Moreover, for a fixed $p \in \mathcal{Z}^+$, with the assumption

$$\sum_{j=-\infty}^{\infty} |j| \alpha^2(j) < \infty \tag{4.2.12}$$

replacing that of $E\{A^4(t)\} < \infty$:

$$n^{1/2}(\tilde{\rho}(1) \ \cdots \ \tilde{\rho}(p))' \xrightarrow{L} N(\mathbf{0}, P_p), \tag{4.2.13}$$

where P_p is the $p \times p$ asymptotic variance-covariance matrix for the serial correlations, with elements given by Bartlett's formula (no 4th cumulant of $A(\cdot)$ is involved). □

Notice that the $n^{1/2}$ in (4.2.11) and (4.2.13) may be replaced by $(n-l)^{1/2}$, where l is any fixed number.

Suppose $d \in \mathcal{N}$, $\{Z(t)\} \in \tilde{I}_d$, $b \in \mathcal{N}$ and $b \geq \max\{d-1, 0\}$. For a fixed $k \in \mathcal{Z}^+$, define the following $k \times (k+b-d)$ matrix

$$G_{bd} = \begin{pmatrix} \frac{g_{bd}(1,1)}{1^{2b}} & \cdots & \frac{g_{bd}(1,1+b-d)}{1^{2b}} & 0 & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & 0 \\ \frac{g_{bd}(k,1)}{k^{2b}} & & \cdots & & & \frac{g_{bd}(k,k+b-d)}{k^{2b}} \end{pmatrix} \qquad (4.2.14a)$$

and then the $k \times (k+b-d+1)$ matrix

$$\dot{G}_{bd} = [(g_{bd}(1,0)/1^{2b} \ \cdots \ g_{bd}(k,0)/k^{2b})' \quad G_{bd}]. \qquad (4.2.14b)$$

Also define

$$\tilde{\gamma}_{bd}(h) = \hat{\gamma}_{bd}(h) - \gamma_{bd}(h), \quad \tilde{\gamma}_{bd} = (\tilde{\gamma}_{bd}(1) \ \cdots \ \tilde{\gamma}_{bd}(k))'. \qquad (4.2.15)$$

In view of first (2.2.24) and (4.2.6), and then (1.4.20) and (4.1.2), we have

$$\tilde{\gamma}_{bd} = \dot{G}_{bd}\tilde{\gamma} + o(n^{-1/2}) \quad a.s., \qquad (4.2.16)$$

where

$$\tilde{\gamma} = (\tilde{\gamma}(0) \ \cdots \ \tilde{\gamma}(k+b-d))'. \qquad (4.2.17)$$

Now $W(t)$ has the form (4.2.8) with i.i.d. stochastic shocks, $A(t)$. So $\{W(t)\}$ is ergodic (see Hannan 1970, p 204), (4.2.8) and (4.2.9) imply $E\{W^2(t)\} < \infty$, and then (see Hannan 1970, p 203), as $n \to \infty$,

$$\hat{\gamma}(0) \to \gamma(0) \quad a.s.. \qquad (4.2.18)$$

Thus, in view of (2.2.24), (4.2.6) and (4.2.10),

$$\frac{\hat{v}_{bd}(h)}{\hat{\gamma}(0)} - \frac{v_{bd}(h)}{\gamma(0)} = \sum_{i=1}^{h+b-d} g_{bd}(h,i)\tilde{\rho}(i) + o(n^{-1/2}) \quad a.s., \ h = 1, \ldots, k. \qquad (4.2.19)$$

Next, define

$$\tilde{\rho}_{bd}(h) = \hat{\rho}_{bd}(h) - \rho_{bd}(h), \quad \tilde{\rho}_{bd} = (\tilde{\rho}_{bd}(1) \ \cdots \ \tilde{\rho}_{bd}(k))'. \qquad (4.2.20)$$

Then, dividing (4.2.19) by h^{2b}, we get

$$\tilde{\rho}_{bd} = G_{bd}\tilde{\rho} + o(n^{-1/2}) \quad a.s., \tag{4.2.21}$$

where

$$\tilde{\rho} = (\tilde{\rho}(1) \ \cdots \ \tilde{\rho}(k+b-d))'. \tag{4.2.22}$$

Theorem 4.2.2 *Suppose $d \in \mathcal{N}$, $\{Z(t)\} \in \tilde{I}_d$, $b \in \mathcal{N}$, $b \geq \max\{d-1, 0\}$, the SPV and SSPV are defined by (4.1.2) and (4.1.4), and the hub series, $\{W(t)\}$, satisfies the conditions in the preliminary theorem. Then, for a fixed $k \in \mathcal{Z}^+$, as $n \to \infty$,*

$$n^{1/2}\tilde{\gamma}_{bd} \xrightarrow{L} N(\mathbf{0}, \dot{G}_{bd}\Gamma_{k+b-d}\dot{G}'_{bd}); \tag{4.2.23}$$

where (to obtain a better approximation when n is not very large) it may be preferable to replace the LHS of this with

$$n^{1/2}\tilde{\tilde{\gamma}}_{bd} \stackrel{def}{=} ((n-l_1)^{1/2}\tilde{\gamma}_{bd}(1) \cdots (n-l_k)^{1/2}\tilde{\gamma}_{bd}(k))'$$

for some fixed constants $l_1, \ldots, l_k$ (in practice, we suggest using $l_h = h$ or $l_h = h+b-1$).

Futher, when $b \geq d$,

$$n^{1/2}\tilde{\rho}_{bd} \xrightarrow{L} N(\mathbf{0}, G_{bd}P_{k+b-d}G'_{bd}); \tag{4.2.24}$$

where it may again be preferable to replace the LHS with

$$n^{1/2}\tilde{\tilde{\rho}}_{bd} \stackrel{def}{=} ((n-l_1)^{1/2}\tilde{\rho}_{bd}(1) \ \cdots \ (n-l_k)^{1/2}\tilde{\rho}_{bd}(k))' \tag{4.2.25}$$

for the fixed constants $l_1, \ldots, l_k$. While, when $b = d-1$,

$$\tilde{\rho}_{b,b+1}(1) = O(1/n) \quad a.s.; \tag{4.2.26}$$

and, if instead of (4.2.20) and (4.2.14a), we put

$$\tilde{\rho}_{b,b+1} = (\tilde{\rho}_{b,b+1}(2) \ \cdots \ \tilde{\rho}_{b,b+1}(k))', \tag{4.2.27}$$

$$G_{b,b+1} = \begin{pmatrix} \frac{g_{b,b+1}(2,1)}{2^{2b}} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ \frac{g_{b,b+1}(k,1)}{k^{2b}} & \cdots & \cdots & \frac{g_{b,b+1}(k,k-1)}{k^{2b}} \end{pmatrix}, \tag{4.2.28}$$

then, again, (4.2.24) holds (with perhaps a corresponding modification to the LHS being preferable).

Proof Using the preliminary theorem, (4.2.23) and (4.2.24) follow directly from (4.2.16) and (4.2.21), respectively. But we have to prove that G_{bd}, defined by (4.2.14a) [or (4.2.28)], is of full rank. [Then, $\dot{G}_{bd}$ defined by (4.1.14b) is of full rank also.] For that, we only need to show

$$g_{bd}(h, h+b-d) \neq 0, \quad h = 1, \quad \dots, \tag{4.2.29}$$

since the last k columns of G_{bd} are lower triangular.

When $b \geq d$, in view of (2.2.25),

$$g_{bd}(h, h+b-d) = 2\pi^h_{bd}(h+b-d)\pi^h_{bd}(0), \quad h = 1, 2, \dots . \tag{4.2.30}$$

From (1.5.15) and (2.2.17), $\pi^h_{bd}(0) = 1$ always holds, and from (2.2.19),

$$\pi^h_{bd}(h+b-d) = \frac{1}{b!}(-1)^{b+1-d}(h+b-1)^{(b)} \neq 0, \quad h = 1, 2, \dots . \tag{4.2.31}$$

So, (4.2.29) is true.

When $b = d-1$, in view of (2.2.26),

$$g_{b,b+1}(h, h-1) = (2 - \delta_{h-1})(b+h-1)^{(b)}/b! > 0, \quad h = 1, 2, \dots . \tag{4.2.32}$$

Again (4.2.29) holds and $G_{b,b+1}$ defined by (4.2.28) is of full rank (so $\dot{G}_{b,b+1}$ is as well). In view of (4.1.13), (4.1.14) and (2.2.25),

$$\begin{aligned} \hat{v}_{b,b+1}(1) &= \frac{g_{b,b+1}(1,0)}{n-b} \sum_{t=b+1}^{n} W^2(t) \\ &= \frac{g_{b,b+1}(1,0)}{n-b}\{(n-m)\hat{\gamma}(0) + (\sum_{t=b+1}^{m} - \sum_{t=m+1}^{b})W^2(t)\}, \end{aligned} \tag{4.2.33}$$

on using (4.1.5). But, in view of (2.2.24), $v_{b,b+1}(1) = g_{b,b+1}(1,0)\gamma(0)$; and then, from the definitions (4.2.20), (4.1.4), (4.1.3), (4.1.2), (1.4.20) and (4.2.33),

$$\tilde{\rho}_{b,b+1}(1) = \frac{g_{b,b+1}(1,0)}{n-b}\{b - m + \frac{1}{\hat{\gamma}(0)}(\sum_{t=b+1}^{m} - \sum_{t=m+1}^{b})W^2(t)\} \tag{4.2.34}$$

and (4.2.26) follows. □

From (4.2.34) we see that $\tilde{\rho}_{b,b+1}(1) = 0$ if we choose $m = b$.

We now turn to the other type of polyvariogram. As mentioned above, we need only consider $b \in \mathcal{Z}^+$ since $\gamma^*_{0d}(h) = \gamma_{0d}(h)/2$. In view of (2.3.40), corresponding to (4.2.14), for a fixed $k \in \mathcal{Z}^+$, define a $k \times (k+b-d)$ matrix,

$$G^*_{bd} = \begin{pmatrix} \frac{g^*_{bd}(1+b,1)}{(1+b)^{2b}} & \cdots & \frac{g^*_{bd}(1+b,1+b-d)}{(1+b)^{2b}} & 0 & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & 0 \\ \frac{g^*_{bd}(k+b,1)}{(k+b)^{2b}} & & \cdots & & & \frac{g^*_{bd}(k+b,k+b-d)}{(k+b)^{2b}} \end{pmatrix}, \qquad (4.2.35a)$$

and then the $k \times (k+b-d+1)$ matrix,

$$\dot{G}^*_{bd} = [(g^*_{bd}(1+b,0)/(1+b)^{2b} \cdots g^*_{bd}(k+b,0)/(k+b)^{2b})' \quad G^*_{bd}]. \qquad (4.2.35b)$$

Also define

$$\tilde{\gamma}^*_{bd}(h) = \hat{\gamma}^*_{bd}(h) - \gamma^*_{bd}(h), \quad \tilde{\gamma}^*_{bd} = (\tilde{\gamma}^*_{bd}(1+b) \cdots \tilde{\gamma}^*_{bd}(k+b))'. \qquad (4.2.36)$$

Previously, we derived a formula for $\gamma^*_{bd}(h)$, (2.3.40), via (2.3.35), by using (1.4.21), (1.4.19) and (1.4.13) [which combine to give (4.1.7)] with (2.2.24). Similarly, from (4.1.8) and (4.2.6) [instead of (4.1.7) and (2.2.24)], we get

$$\hat{\gamma}^*_{bd}(h) = \frac{(-1)^{b+1}}{h^{2b}} \sum_{i=0}^{h-d} g^*_{bd}(h,i)\hat{\gamma}(i) + o(n^{-1/2}) \quad a.s., \ h = 1+b, \dots, k+b, \qquad (4.2.37)$$

and hence

$$\tilde{\gamma}^*_{bd} = (-1)^{b+1} \dot{G}^*_{bd} \tilde{\gamma} + o(n^{-1/2}) \quad a.s.\,, \qquad (4.2.38)$$

where $\tilde{\gamma}$ is given by (4.2.17). Using (4.1.9) and (4.1.10), define

$$\tilde{\rho}^*_{bd}(h) = \hat{\rho}^*_{bd}(h) - \rho^*_{bd}(h), \quad \tilde{\rho}^*_{bd} = (\tilde{\rho}^*_{bd}(1+b) \cdots \tilde{\rho}^*_{bd}(k+b))'. \qquad (4.2.39)$$

Then from (4.2.18), (4.2.37) and (2.3.40), we have

$$\tilde{\rho}^*_{bd} = (-1)^{b+1} G^*_{bd} \tilde{\rho} + o(n^{-1/2}) \quad a.s.\,, \qquad (4.2.40)$$

where $\tilde{\rho}$ is given by (4.2.22).

Theorem 4.2.3 *Under the same conditions as in Theorem 4.2.2,*

$$n^{1/2} \tilde{\gamma}^*_{bd} \xrightarrow{L} N(0,\ \dot{G}^*_{bd}\ \Gamma_{k+b-d}\ \dot{G}^{*\prime}_{bd}). \qquad (4.2.41)$$

When $b \geq d$,

$$n^{1/2}\tilde{\rho}^*_{bd} \xrightarrow{L} N(0,\ G^*_{bd}\ P_{k+b-d}\ G^{*\prime}_{bd}). \tag{4.2.42}$$

When $b = d-1$, then

$$\tilde{\rho}^*_{b,b+1}(1+b) = O(1/n) \quad a.s.. \tag{4.2.43}$$

Instead of (4.2.39) and (4.2.35a), write:

$$\tilde{\rho}^*_{b,b+1} = (\tilde{\rho}^*_{b,b+1}(2+b) \quad \cdots \quad \tilde{\rho}^*_{b,b+1}(k+b))', \tag{4.2.44}$$

$$G^*_{b,b+1} = \begin{pmatrix} \frac{g^*_{b,b+1}(2+b,1)}{(2+b)^{2b}} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ \frac{g^*_{b,b+1}(k+b,1)}{(k+b)^{2b}} & \cdots & \cdots & \frac{g^*_{b,b+1}(k+b,k)}{(k+b)^{2b}} \end{pmatrix}. \tag{4.2.45}$$

Then, (4.2.42) again holds.

Proof The argument is similar to that for the previous theorem. This time, we need G^*_{bd} to be of full rank; for which it is sufficient to show that

$$g^*_{bd}(h, h-d) \neq 0, \quad h = b+1,\ b+2, \dots\ . \tag{4.2.46}$$

Now, for $h > b$, (2.3.34) gives

$$g^*_{bd}(h, h-d) = (-1)^{b+1}\frac{bh}{2}\frac{(b-1)!}{h^{(b+1)}}g_{bd}(h-b, h-d) \neq 0,$$

on using (4.2.29), which yields (4.2.46).

From (4.1.7) and (4.1.8), we have

$$\rho^*_{b,b+1}(1+b) = \frac{\gamma^*_{b,b+1}(1+b)}{\gamma(0)} = \frac{v_{b,b+1}(1)}{2(b+1)^{2b}\gamma(0)} \tag{4.2.47a}$$

and

$$\hat{\rho}^*_{b,b+1}(1+b) = \frac{\hat{\gamma}^*_{b,b+1}(1+b)}{\hat{\gamma}(0)} = \frac{\hat{v}_{b,b+1}(1)}{2(b+1)^{2b}\hat{\gamma}(0)}. \tag{4.2.47b}$$

So, referring to (4.2.39) and (4.2.33), and again noticing that $v_{b,b+1}(1) = g_{b,b+1}(1,0)\gamma(0)$, we get (4.2.43). □

Remark In the case of the γ^*_b-polyvariogram, we do not suggest using a modification like (4.2.25) in practice. For the γ_b-polyvariogram, the $\hat{v}_b(h)$ in (4.1.2) is an average of $n-h-b+1$ terms; and, intuitively, when n is not very large, using the $n^{1/2}\tilde{\tilde{\rho}}_{bd}$ of (4.2.25) and $n^{1/2}\tilde{\tilde{\gamma}}_{bd}$ may give better normal approximations. But, for the γ_b^* family, (4.1.8) shows that $\hat{\gamma}_b^*(h)$ is already a weighted average of the $\hat{v}_b(j)$, $j=1,\ldots,h-b$, with weights that decrease as j increases. So a promising modification analogous to (4.2.25) (or its companion modification) is not obvious. Perhaps the ready availability of a beneficial modification is one practical advantage that $\hat{\gamma}_b$ has over $\hat{\gamma}_b^*$, although the asymptotic properties of both estimators are similar.

4.2.3 The almost sure convergence of the SPVs and SSPVs

To discuss the almost sure convergence, we need different assumptions on the hub series, $\{W(t)\}$, [cf (4.2.8) and (4.2.9)]. On the one hand, rather than assuming the i.i.d. innovation shocks, $\{A(t)\}$, we make the following less stringent assumptions. We suppose that the shocks are a series of martingale differences, $\{M(t)\}$, i.e.,

$$E\{M(t) \mid \mathcal{F}_{t-1}\} = 0 \quad a.s. \; ; \tag{4.2.48}$$

and, futher, that these $M(t)$ are strictly stationary and ergodic, and satisfy

$$E\{M^2(t) \mid \mathcal{F}_{t-1}\} = \sigma^2 \quad a.s., \qquad E\{M^4(t)\} < \infty, \tag{4.2.49}$$

where $\mathcal{F}_t$ is the σ-algebra generated by $\{M(t), M(t-1), \ldots\}$. On the other hand, we restrict the general linear series (4.2.8) to being a linear series

$$W(t) = \sum_{j=0}^{\infty} \alpha(j) M(t-j), \qquad \sum_{j=0}^{\infty} \mid \alpha_j \mid < \infty. \tag{4.2.50}$$

Then there are some results (see An et al. 1982) about the almost sure convergence of $\hat{\gamma}(i)$ defined by (4.1.6). Among them, say: if, in addition to (4.2.49) and (4.2.50), $\{W(t)\}$ is an ARMA series, and if $p(n) = O\{(\log n)^a\}$ for some a, $0 < a < \infty$ [i.e., $p(n) \to \infty$, as $n \to \infty$, but at the much slower rate indicated]; then

$$\max_{0 \le i \le p(n)} \mid \tilde{\gamma}(i) \mid = O(q_n) \quad a.s., \tag{4.2.51}$$

where $\tilde{\gamma}(i) = \hat{\gamma}(i) - \gamma(i)$, and here (and in the remainder of this thesis), for brevity, we denote

$$q_n = (n^{-1} \log \log n)^{1/2}. \tag{4.2.52}$$

A corresponding result holds for $\tilde{\rho}(i) = \hat{\rho}(i) - \rho(i)$ under even weaker conditions (see Hannan and Kavalieris 1983):

$$\max_{0\leq i\leq p(n)} \mid \tilde{\rho}(i) \mid = O(q_n) \quad a.s. \quad . \tag{4.2.53}$$

As we have pointed out, when i is restricted to a finite range, then the asymptotic properties of $\hat{\gamma}(i)$ are the same irrespective of whether $\hat{\gamma}(i)$ is defined by (4.1.5) or by (4.1.6). So (4.2.51) and (4.2.53) also hold if $\hat{\gamma}(i)$ is defined by (4.1.5). Further, conditions (4.2.48) to (4.2.50) imply the conditions in Lemma 4.2.2, and hence (4.2.6) holds. Thus, we may easily show that [under the conditions which ensure (4.2.51)], for a fixed $k \in \mathcal{Z}^+$,

$$\max_{1\leq h\leq k} \tilde{\gamma}_{bd}(h) = O(q_n), \qquad \max_{1\leq h\leq k} \tilde{\rho}_{bd}(h) = O(q_n) \quad a.s. \ , \tag{4.2.54}$$

$$\max_{1+b\leq h\leq k+b} \tilde{\gamma}^{*}_{bd}(h) = O(q_n), \qquad \max_{1+b\leq h\leq k+b} \tilde{\rho}^{*}_{bd}(h) = O(q_n) \quad a.s. \ . \tag{4.2.55}$$

4.3 Determining the Degree of Differencing for Integrated White Noise

4.3.1 Setting up hypotheses

We have introduced a graphical procedure for determining the degree of differencing. This procedure is simple and convenient to use when the graphical features obseved are clear-cut reflections of their theoretical ideals. However, when the graphical features are ambiguous, further investigation is needed. Theorems concerning the asymptotic properties of the SPVs or the SSPVs, some established in the last section and others to be subsequently established, provide more sensitive tools for this purpose.

Suppose that, from inspection of some sequence of the SPVs or the SSPVs, we have formed a somewhat subjective impression as to the geometry of the asymptotes of the corresponding theoretical quantities and have made an initial tentative judgement as to the true value of d. Then this judgement provides us with a null hypothesis, $H_0^{(d)} : \{Z(t)\} \in \tilde{I}_d$.

However, the graphical features may prove ambiguous. For example, $\hat{\gamma}_2(h)$ could level out and $\hat{\gamma}_0(h)$ have a marked upward sloping linear trend, but it might be unclear as to whether $\hat{\gamma}_1(h)$ levels out or has a slightly positive sloping asymptote. Then the possibilities for d could be 1 or 2; and we may first test $H_0^{(2)}$ and next test $H_0^{(1)}$, if $H_0^{(2)}$ were rejected.

We may also totally abandon the graphical procedure and set up hypotheses to determine d in the following way. First guess a safe upper bound for d, d_0 say, and test $H_0^{(d_0)}$. As already mentioned, $d_0 = 3$ or 4 should usually be reasonable in practice. If $H_0^{(d_0)}$ is rejected by a single testing procedure, next test $H_0^{(d_0-1)}$, and so on down until, for some $\hat{d}$, $H_0^{(\hat{d})}$ is accepted by all testing procedures. Then $\hat{d}$ is our estimate of d. As one can not rely on a hypothesis being accepted by a classical Neyman-Pearson test (based on the distribution), any battery of testing procedures which accepts $H_0^{(\hat{d})}$ should include one which is based on almost sure convergence results — since we evaluate the effectiveness of such a procedure by its consistency and not by the probabilities of the two types of errors. We will explain this more precisely in subsection 4.3.4.

For $H_0^{(d)}$, from the previous section we see that our test statistics can be chosen as $\tilde{\gamma}_{bd}(h) = \hat{\gamma}_{bd}(h) - \gamma_{bd}(h)$, $\tilde{\rho}_{bd} = \hat{\rho}_{bd}(h) - \rho_{bd}(h)$ with suitable modification ($b \geq \max\{d-1, 0\}$), or we can use $\tilde{\gamma}^*_{bd}$ or $\tilde{\rho}^*_{bd}$. Here, $\hat{\gamma}_{bd}(h)$ say, is merely a specific formula for $\hat{\gamma}_b(h)$ [under the assumption of $\{Z(t)\} \in \tilde{I}_d$] and, in practice, we then use $\hat{\gamma}_b(h)$ given by (4.1.2) which is only a function of b, h and the data. But $\gamma_{bd}(h)$ is a function of b, d, h and also the $\gamma(i)$, the autocovariances of the hub series, which are unknown. So we can not obtain $\tilde{\gamma}_{bd}(h)$, and hence the results of Section 4.2 are not very helpful for hypothesis testing, in the general situation, although they are useful for obtaining confidence intervals. Fortunately, when $\{Z(t)\}$ is an integrated white noise [$\rho(0) = 1$ and $\rho(j) = 0,\ j \neq 0$], from (2.2.33), (2.3.40) and (4.1.3), we have

$$\rho_{bd}(h) = g_{bd}(h, 0)/h^{2b}, \quad \rho^*_{bd}(h) = g^*_{bd}(h, 0)/h^{2b}, \tag{4.3.1}$$

and hence $\tilde{\rho}_{bd}(h) = \hat{\rho}_b(h) - \rho_{bd}(h)$ and $\tilde{\rho}^*_{bd}(h) = \hat{\rho}^*_b(h) - \rho_{bd}(h)$ can be obtained from the data and used as test statistics.

The formulae defining $\hat{\rho}_b(h)$ and $\hat{\rho}_b^*(h)$ were given in Section 4.1. The general expressions for $g_{bd}(h,0)$ and $g_{bd}^*(h,0)$ are (2.2.25) and (2.3.34), which were obtained from the IPIV. We may also obtain $g_{bd}^*(h,0)$ from the solution of the difference equation (2.3.6), as was done in subsection 2.3.2. The special formulae for low orders ($b = 0$, 1 and 2) can be found from (2.2.32) and (2.3.39), where the $g_{bd}(h,0)$ or $g_{bd}^*(h,0)$ are the coefficients of $\gamma(0)$ in these formulae.

4.3.2 The underlying theorems

In Theorems 4.2.2 and 4.2.3, if $W(t) = A(t)$ [see (4.2.9)], then P_{k+b-d} in (4.2.24) and (4.2.42) becomes the unit matrix and then the asymptotic distributions of $\tilde{\rho}_{bd}$ and $\tilde{\rho}_{bd}^*$ are clear from these theorems. These distributions can be used to derive testing procedures, but this subsection will provide a theorem giving the same results under weaker conditions: $W(t) = M(t)$, where the $M(t)$ are defined by the material from (4.2.48) to (4.2.49) inclusive. Moreover, in this situation, although (4.2.54) and (4.2.55) are certainly true, we can actually prove more precise results from which more delicate testing procedures can be designed.

Lemma 4.3.1 *Suppose the $M(t)$ are strictly stationary, ergodic martingale differences given by (4.2.48) and which also satisfy (4.2.49). Then, for any fixed $m \in \mathcal{N}$, as $n \to \infty$,*

$$\hat{\sigma}^2 = \frac{1}{n-m} \sum_{t=m+1}^{n} M^2(t) \to \sigma^2 \quad a.s.\,; \tag{..3.2}$$

and, for any k' constants, say $c_1, \ldots, c_{k'}$ with $c^2 = \sum_{i=1}^{k'} c_i^2 > 0$,

$$\left(\frac{n}{c^2\hat{\sigma}^4}\right)^{1/2} \sum_{i=1}^{k'} c_i \left\{\frac{1}{n-m} \sum_{t=m+1}^{n} M(t)M(t-i)\right\} \xrightarrow{L} N(0,1) \tag{4.3.3}$$

and

$$\limsup_{n\to\infty} \left(\frac{n}{2c^2\hat{\sigma}^4 \log\log n}\right)^{1/2} \left| \sum_{i=1}^{k'} c_i \left\{\frac{1}{n-m} \sum_{t=m+1}^{n} M(t)M(t-i)\right\} \right| = 1 \quad a.s.\,. \tag{4.3.4}$$

Proof Due to the conditions on the $M(t)$, both $M^2(t) - \sigma^2$ and $M(t-i)M(t-j)$, $i \neq j$, are square-integrable martingale differences which are also strictly stationary and ergodic (see Doob 1953, p 458). So the law of the iterated logarithm (LIL) holds (Stout 1970), which implies that

$$\frac{1}{n-m}\sum_{t=m+1}^{n}\{M^2(t)-\sigma^2\}=O(q_n) \quad a.s., \tag{4.3.5}$$

$$\frac{1}{n-m}\sum_{t=m+1}^{n}M(t-i)M(t-j)=O(q_n) \quad a.s., \quad i\neq j, \tag{4.3.6}$$

where, again, $q_n=(n^{-1}\log\log n)^{1/2}$; and (4.3.2) follows from (4.3.5). Define

$$X(t)=M(t)\{\sum_{i=1}^{k'}c_iM(t-i)\}. \tag{4.3.7}$$

For the same reasons as above, the $X(t)$ are strictly stationary, ergodic and square-integrable martingale differences. Also it is easy to show that

$$E\{X^2(t)\}=c^2\sigma^4. \tag{4.3.8}$$

From (4.3.5), (4.3.6) and (4.3.8), we have

$$\sum_{t=m+1}^{n}E\{X^2(t)\mid\mathcal{F}_{t-1}\}/E\{\sum_{t=m+1}^{n}X^2(t)\}\to 1 \quad a.s.\,. \tag{4.3.9}$$

Then the central limit theorem (CLT) holds (see Hall and Heyde 1980, p 51) for $\{X(t)\}$, and (4.3.3) follows from (4.3.2), (4.3.7) and (4.3.8). Also, from Stout (1970), the LIL holds for $\{X(t)\}$ and (4.3.4) follows in the same way. □

Denote

$$C_{bd}=G_{bd}G'_{bd}, \qquad C^*_{bd}=G^*_{bd}G^{*\prime}_{bd}, \tag{4.3.10}$$

where G_{bd} is defined by (4.2.14a) ($b\geq d$) or (4.2.28) ($b=d+1$) and G^*_{bd} is defined by (4.2.35a) ($b\geq d$) or (4.2.45) ($b=d+1$). Note that [cf (4.2.14a) and (4.2.35a)]

$$g_{bd}(h,i)=0, \quad \text{when } i>h+b-d; \quad g^*_{bd}(h,i)=0, \quad \text{when } i>h-d; \tag{4.3.11}$$

and let

$$\delta=0 \quad \text{if} \quad b\geq d, \qquad \delta=1 \quad \text{if} \quad b=d-1. \tag{4.3.12}$$

Then the $(h-\delta,j-\delta)$-th elements of C_{bd} are

$$c_{bd}(h,j)=\sum_{i=1}^{k+b-d}g_{bd}(h,i)g_{bd}(j,i)/(hj)^{2b} \tag{4.3.13}$$

and the $(h-b-\delta, j-b-\delta)$-th elements of C^*_{bd} are

$$c^*_{bd}(h,j) = \sum_{i=1}^{k-d} g^*_{bd}(h,i) g^*_{bd}(j,i)/(hj)^{2b}. \tag{4.3.14}$$

For $b = 0, 1$, and 2, the special values of $g_{bd}(h,i)$ and $g^*_{bd}(h,i)$ are given in (3.2.32) and (2.3.39), where they are the coefficients of $\gamma(i)$. It was shown in the proof of Theorems 4.2.2 and 4.2.3 that G_{bd} and G^*_{bd} are of full rank, so

$$C_{bd} > 0, \qquad C^*_{bd} > 0. \tag{4.3.15}$$

Combining (4.2.20) and (4.2.27), or (4.2.39) and (4.2.44), using (4.3.12), we may write

$$\tilde{\boldsymbol{\rho}}_{bd} = (\tilde{\rho}_{bd}(1+\delta) \ \cdots \ \tilde{\rho}_{bd}(k))', \quad \tilde{\boldsymbol{\rho}}^*_{bd} = (\tilde{\rho}^*_{bd}(1+\delta+b) \ \cdots \ \tilde{\rho}^*_{bd}(k+b))'. \tag{4.3.16}$$

Theorem 4.3.1 *Suppose $d \in \mathcal{N}$, the $M(t) = \nabla^d Z(t)$ are strictly stationary, ergodic martingale differences given by (4.2.48) and satisfying (4.2.49), and $b \in \mathcal{N}$, $b \geq$* $\max\{d-1, 0\}$. *Then, for a fixed $k \in \mathcal{Z}^+$, as $n \to \infty$, the CLT holds for the SSPV:*

$$n^{1/2} \tilde{\boldsymbol{\rho}}_{bd} \xrightarrow{L} N(\mathbf{0}, C_{bd}), \tag{4.3.17}$$

$$n^{1/2} \tilde{\boldsymbol{\rho}}^*_{bd} \xrightarrow{L} N(\mathbf{0}, C^*_{bd}). \tag{4.3.18}$$

Proof We only give the proof of (4.3.17); (4.3.18) can be proved similarly. Notice that (4.2.21) is true provided that (4.2.18) holds and the conditions on $W(t)$ in Theorem 4.2.1 are valid — all of which is so for $W(t) = M(t)$ [see (4.2.48), (4.2.49) and (4.3.2)]. So, for any constant vector, $\boldsymbol{\lambda} = (\lambda_{1+\delta} \ \cdots \ \lambda_k)'$, in view of (4.2.21) and because $\tilde{\rho}(i) = \hat{\rho}(i)$ for $i \neq 0$, and $\hat{\gamma}(0) = \hat{\sigma}^2$ [since $W(t) = M(t)$], we have

$$\begin{aligned}\boldsymbol{\lambda}' n^{1/2} \tilde{\boldsymbol{\rho}}_{bd} &= n^{1/2} \sum_{h=1+\delta}^{k} \lambda_h \sum_{i=1}^{h+b-d} \{\frac{g_{bd}(h,i)}{h^{2b}} \frac{\hat{\gamma}(i)}{\hat{\gamma}(0)}\} + o(1) \\ &= (\frac{n}{\hat{\sigma}^4})^{1/2} \sum_{i=1}^{k+b-d} \{ \sum_{h=1+\delta}^{k} \lambda_h \frac{g_{bd}(h,i)}{h^{2b}} \} \hat{\gamma}(i) + o(1) \quad a.s.\,.\end{aligned} \tag{4.3.19}$$

Then using (4.3.3) and (4.3.13), we get

$$\boldsymbol{\lambda}' n^{1/2} \tilde{\boldsymbol{\rho}}_{bd} \xrightarrow{L} N(0, c^2), \qquad c^2 = \boldsymbol{\lambda}' C_{bd} \boldsymbol{\lambda}, \tag{4.3.20}$$

where, $C_{bd} > 0$. By the Cramér-Wold device (see Billingsley 1968, p 49), (4.3.17) follows. □

Theorem 4.3.2 *Under the same conditions as in Theorem 4.3.1, for a fixed $h \in \mathcal{Z}^+$ ($h \neq 1$ if $b = d-1$), as $n \to \infty$, the LIL holds for the SSPV:*

$$\limsup_{n\to\infty}[n/\{2c_{bd}(h,h)\log\log n\}]^{1/2} \mid \tilde{\rho}_{bd}(h) \mid = 1 \quad a.s. , \tag{4.3.21}$$

$$\limsup_{n\to\infty}[n/\{2c^*_{bd}(h,h)\log\log n\}]^{1/2} \mid \tilde{\rho}^*_{bd}(h+b) \mid = 1 \quad a.s. . \tag{4.3.22}$$

Proof First, $c_{bd}(h,h) > 0$ holds, since $C_{bd} > 0$. In view of (4.2.21), similarly to (4.3.19), the "argument of lim sup" in the LHS of (4.3.21) is

$$\{\frac{n}{2c_{bd}(h,h)\hat{\sigma}^4\log\log n}\}^{1/2}\sum_{i=1}^{h+b-d}\frac{g_{bd}(h,i)}{h^{2b}}\hat{\gamma}(i) + o(1) \quad a.s. . \tag{4.3.23}$$

Then, from (4.3.4) and (4.3.13), (4.3.21) follows. (4.3.22) can be similarly proved. □

In the above theorems, n can be replaced by $n-m$, given any fixed number m. [$m = h$ or $m = h+b-1$ are recommended for (4.3.21); for (4.3.17), instead of using $\tilde{\rho}_{bd}$, $\tilde{\tilde{\rho}}_{bd}$ defined by (4.2.25) is preferable].

4.3.3 Tests based on the CLT results

Using result (4.3.17) or (4.3.18) to test $H_0^{(d)}$: $\{Z(t)\} \in \tilde{I}_d$ is classical. Say, (4.3.17) gives $(n-h)^{1/2}\tilde{\rho}_{bd}(h) \sim N(0, c_{bd}(h,h))$ asymptotically. So the test procedure for an individual $h \in \{1+\delta, \ldots, k\}$ is standard. We could then reject $H_0^{(d)}$ if one of these individual tests rejected $H_0^{(d)}$.

However, should we do that, the actual significance of such a battery of tests will be difficult to determine. What we would prefer are overall tests — that is "portmanteau" tests, in the spirit of those proposed by Box and Jenkins (1976). From (4.3.17), (4.3.18) and (4.3.15), asymptotically,

$$n\tilde{\rho}'_{bd}C_{bd}^{-1}\tilde{\rho}_{bd} \sim \chi^2_{k-\delta}, \quad n\tilde{\rho}^{*\prime}_{bd}C_{bd}^{*-1}\tilde{\rho}^*_{bd} \sim \chi^2_{k-\delta}. \tag{4.3.24}$$

The test produre is well-known.

4.3.4 Tests based on the LIL results

We only derive the procedures based on (4.3.21), those based on (4.3.22) may be similarly obtained.

For a fixed $k \in \mathcal{Z}^+$, $1+\delta \le h \le k$, write

$$R_{bd}(h) = \mid \tilde{\rho}_{bd}(h) \mid / \{2c_{bd}(h,h)\}^{1/2}, \tag{4.3.25}$$

$$R_{bd} = \max_{1+\delta \le h \le k} R_{bd}(h) \tag{4.3.26}$$

and, again, $q_n = (n^{-1}\log\log n)^{1/2}$. Then, in view of (4.3.21), if $H_0^{(d)} : \{Z(t)\} \in \tilde{I}_d$ is true; we have, for $h \in \{1+\delta, \ldots, k\}$, $\limsup_{n\to\infty} R_{bd}(h)/q_n = 1$ *a.s.* . Since R_{bd}, defined by (4.3.26), depends only on a finite number of $R_{bd}(h)$,

$$\limsup_{n\to\infty} R_{bd}/q_n = 1 \quad a.s. \, . \tag{4.3.27}$$

Take a sequence of positive numbers p_n, such that

$$\lim_{n\to\infty} p_n = 0; \; \lim_{n\to\infty} \; p_n/q_n > 1 \text{ (maybe } \infty). \tag{4.3.28}$$

Such a sequence could be, say,

$$p_n = (1+\varepsilon)q_n, \qquad 0 < \varepsilon < 1, \tag{4.3.29}$$

then

$$\limsup_{n\to\infty} \; R_{bd}/p_n = 1/(1+\varepsilon) < 1 \qquad a.s. \, . \tag{4.3.30}$$

Such a sequence could also be, say,

$$p_n = (n^{-1}\log n)^{1/2}, \tag{4.3.31}$$

then

$$\limsup_{n\to\infty} \; R_{bd}/p_n = 0 \quad a.s. \quad . \tag{4.3.32}$$

Suppose $H_0^{(d)}$ is false; say, in fact, $\{Z(t)\} \in \tilde{I}_{d'}, d' \neq d$. If $b \ge d'-1$ and $b \ge d-1$, then $\rho_{bd'}(h)$ and $\rho_{bd}(h)$ exist and

$$\max_{1+\delta \le h \le k} \mid \rho_{bd'}(h) - \rho_{bd}(h) \mid > 0. \tag{4.3.33}$$

Notice again, in the definition of $\tilde{\rho}_{bd}(h)$, (4.2.20), $\hat{\rho}_{bd}(h)$ is only an expression for $\hat{\rho}_b(h)$ under the assumption that $\{Z(t)\} \in \tilde{I}_d$; but the values which $\hat{\rho}_b(h)$ takes do not depend on d. When $\{Z(t)\} \in \tilde{I}_{d'}$, $\hat{\rho}_b(h)$ should be expressed as $\hat{\rho}_{bd'}(h)$, and then

$$\tilde{\rho}_{bd}(h) = \hat{\rho}_b(h) - \rho_{bd}(h) = \tilde{\rho}_{bd'}(h) + (\rho_{bd'}(h) - \rho_{bd}(h)), \tag{4.3.34}$$

and hence

$$\frac{|\rho_{bd'}(h)-\rho_{bd}(h)|}{\{2c_{bd}(h,h)\}^{1/2}p_n}-\{\frac{c_{bd'}(h,h)}{c_{bd}(h,h)}\}^{1/2}\frac{R_{bd'}(h)}{p_n}\leq\frac{R_{bd}(h)}{p_n}. \qquad (4.3.35)$$

From (4.3.30) or (4.3.32), with $R_{bd}(h)$ being replaced by $R_{bd'}(h)$ (since now the true degree is d'), the subtracted term in the LHS of (4.3.35) is less than 1 and, from (4.3.28) and (4.3.33), the term it is subtracted from converges to ∞, so

$$\lim_{n\to\infty} R_{bd}/p_n=\infty \qquad a.s.. \qquad (4.3.36)$$

Comparing (4.3.36) (when $H_0^{(d)}$ is false) with (4.3.30) or (4.3.32) (when $H_0^{(d)}$ is true), we see substantially distinct asymptotic behaviours of R_{bd}/p_n for these two possible cases. Choose a c, $=1$ say, as a critical value. Then, when $R_{bd}/p_n<c$, we accept $H_0^{(d)}$, otherwise we reject $H_0^{(d)}$.

According to (4.3.36) and (4.3.30) or (4.3.32), the above decision is right almost surely (i.e., with probability 1) provided that n is sufficiently large. We call such a procedure *consistent.*

In fact, if we use the p_n of (4.3.29), then we may take any c, $1/(1+\varepsilon)<c<\infty$, as a critical value; while, if we use the p_n of (4.3.31), any c, $0<c<\infty$, may be used -- the test procedures resulting from such choices of c being consistent. However, in practice, n is finite and, for series with less than a few hundred terms, an informed choice of p_n and c is crucial. Our suggestion of $c=1$, or thereabouts, for the p_n of (4.3.29) or (4.3.31) seems reasonable according to (4.3.21).

If we use (4.3.29) with $c=1$, the choice of ε is still crucial. We may make the test more sensitive for detecting a false null hypothesis (so that an acceptance of H_0 becomes more reliable) by choosing an appropriate small ε (simulations are needed for such a choice, giving ε dependent on b,d,k and n). Compare (4.3.29) and (4.3.31) — two different choices of p_n. When the chosen value of $\varepsilon>0$ is less than 0.5, say, $(n^{-1}\log n)^{1/2}>(1+\varepsilon)(n^{-1}\log\log n)^{1/2}$ holds for all $n>e$ and hence

$$\frac{R_{bd}}{(n^{-1}\log n)^{1/2}}<\frac{R_{bd}}{(1+\varepsilon)(n^{-1}\log\log n)^{1/2}} \qquad (4.3.37)$$

(where both sides of the inequality are special cases of R_{bd}/p_n). Evidently, using $p_n=(n^{-1}\log n)^{1/2}$, it is easier to accept a false null hypothesis than when using

$p_n = (1+\varepsilon)(n^{-1}\log\log n)^{1/2}$. But due to (4.3.36), when n is not too small, the RHS of (4.3.37) is quite likely to be larger than $c = 1$, and then the false null hypothesis is rejected. Also it is clear that using $p_n = (n^{-1}\log n)^{1/2}$, it is harder to reject a true null hypothesis than when using $p_n = (1+\varepsilon)(n^{-1}\log\log n)^{1/2}$.

We have indicated how to apply (4.3.21); (4.3.22) can be applied similarly. Both results are much more useful than results such as (4.2.54) or (4.2.55), since they may suggest special choices for c which (4.2.54) and (4.2.55) can not do.

4.3.5 Intra-polyvariogram correlation structure

In (4.3.13), we gave the asymptotic covariance, $c_{bd}(h,j)$, for the SSPVs, $\hat{\rho}_{bd}(h)$ and $\hat{\rho}_{bd}(j)$. Similarly, (4.3.14) gave $c^*_{bd}(h,j)$ for $\hat{\rho}^*_{bd}(h)$ and $\hat{\rho}^*_{bd}(j)$. We also proved the covariance matrix of $\tilde{\rho}_{bd}$ and $\tilde{\rho}^*_{bd}$ [see (4.3.16)] is positive definite [see (4.3.15)]. Now define the asymptotic correlation coefficients as

$$r_{bd}(h,j) = c_{bd}(h,j)/\{c_{bd}(h,h)c_{bd}(j,j)\}^{1/2}, \quad 1+\delta \le h \le j \le k, \qquad (4.3.38)$$

and

$$r^*_{bd}(h,j) = c^*_{bd}(h,j)/\{c^*_{bd}(h,h)c^*_{bd}(j,j)\}^{1/2}, \quad 1+\delta+b \le h \le j \le k+b. \qquad (4.3.39)$$

For example, from (2.2.32), (4.3.13) gives $c_{00}(h,j) = 4\delta_{h-j}$ (δ_i is the Kronecker delta function) and $c_{01}(h,j) = 2(h-1)h(3j-h-1)/3$. Then, $r_{00}(h,j) = \delta_{h-j}$ and

$$r_{01}(h,j) = (3j-h-1)\{\frac{(h-1)h}{(2h-1)(j-1)j(2j-1)}\}^{1/2}. \qquad (4.3.40a)$$

Similarly,

$$r_{10}(h,j) = \begin{cases} 3/\sqrt{17} & (h=1,\ j=2) \\ 4/[\{1+1/(j^2+j)\}\sqrt{17}] & (h=1,\ j>2) \\ 1-3j^2/(j^4+j^2+1) & (h>1,\ j=h+1) \\ 1/[\{1+1/(h^2+h)\}\{1+1/(j^2+j)\}] & (h>1,\ j>h+1) \end{cases} \qquad (4.3.40b)$$

$$r_{11}(h,j) = \{\frac{h(h+1)(2h+1)}{j(j+1)(2j+1)}\}^{1/2} \qquad (4.3.40c)$$

$r_{12}(h,j) =$

$$[3(h^2-4)(h+3)(2h+1)+7j(j+1)\{5(3h+2)(2j+1)-6(h+2)(2h+1)\}]$$
$$\times\{\frac{h(h^2-1)}{4(2h+1)(33h^3+29h^2-13h-18)j(j^2-1)(2j+1)(33j^3+29j^2-13j-18)}\}^{1/2}. \tag{4.3.40d}$$

From (4.3.40a), we find that, for h starting from 2, $r_{01}(h,h+1)$ is greater than 0.89 and converges to 1 as h increases. The other $r_{bd}(h,h+1)$, from (4.3.40), have similar properties. This demonstrates that, apart from the $(b,d)=(0,0)$ case, the SSPVs have adjacent values that eventually are highly correlated. The high correlation mainly starts from early h. Among the three cases with $b=1$, only $r_{11}(h,h+1)$ starts off from less than 0.72 (and does not exceed 0.9 until h exceeds 13).

For $r^*_{bd}(h,h+1)$, all four cases with $b=0$ or 1 [not including $(b,d)=(0,0)$] have high values (the lowest starting value exceeding 0.89), and all tend to 1 as h increases. Specifically, the formulae for these four cases are:

$$r^*_{01}(h,j) = (3j-h-1)\{\frac{(h-1)h}{(2h-1)(j-1)j(2j-1)}\}^{1/2} \tag{4.3.41a}$$

$$r^*_{10}(h,j) = \{(1+1/h^4)(1+1/j^4)\}^{-1/2} \tag{4.3.41b}$$

$$r^*_{11}(h,j) = (3j-h-1)\{\frac{(h-1)h}{(2h-1)(j-1)j(2j-1)}\}^{1/2} \tag{4.3.41c}$$

$$r^*_{12}(h,j) = [35(j^3-j)-(h+2)\{7j(3j-h)+h^2-9)]$$
$$\times\{\frac{(h-2)(h^2-1)j^3}{4h^3(2h-1)(5h^2-5h-9)(j-2)(j^2-1)(2j-1)(5j^2-5j-9)}\}^{1/2}. \tag{4.3.41d}$$

This high intra-polyvariogram correlation structure indicates that, except for the (0,0) cases, the SSPVs and hence the SPVs will eventually be very smooth; and, for all but one of them, this smoothness starts very close to the beginning.

4.4 Determining the Degree of Differencing for Integrated ARMA Series

4.4.1 A model transformation

As we explained in the last section, the scaled polyvariograms, $\rho_{bd}(h)$ and $\rho^{*}_{bd}(h)$, can only be obtained for the integrated white noise case [see (4.3.1)]. Otherwise, these scaled polyvariograms depend on the unknown autocovariance function, $\gamma(i)$, of the hub series. But the integrated white noise model is not general enough to meet the modelling requirements of practice. What we would prefer is the ARIMA model which has been widely used by practitioners since the work of Box and Jenkins (1976). However, as the hub (ARMA) series ACVF will not be known, we can no longer set up hypotheses using the actual SSPV like we did in the last section.

A way to overcome this difficulty is to make a model transformation. Consider an ARIMA (p,d,q) series as given by (1.1.1) and (1.1.2); but, for greater theoretical generality, rather than (1.1.2) we consider the hub series satisfying

$$\phi(B)W(t)=\theta(B)M(t), \quad t=\ldots,-1,0,1,\ldots, \tag{4.4.1}$$

where $\phi(B)$ and $\theta(B)$ are relatively prime and satisfy (1.1.3), and the $M(t)$ are defined by the material from (4.2.48) to (4.2.49) inclusive, which gives a more general shock series than an i.i.d. series, $\{A(t)\}$. Then (1.1.4) can be extended to

$$\phi(B)\nabla^{d}Z(t)=\theta(B)M(t), \quad t=p+d,\ p+d+1,\ldots\ . \tag{4.4.2}$$

Here, like (1.1.4), we emphasize the legitimate range of t by combining (1.1.1) with (4.4.1). For $0\le t<p+d$, although the $Z(t)$ are properly defined, (4.4.2) does not hold.

Suppose a series $\{X(t)\}$ (starting from a suitable time) satisfies

$$\phi(B)Z(t)=\theta(B)X(t), \quad t=p,\ p+1,\ldots\ . \tag{4.4.3}$$

Then, differencing both sides d times, we get $\phi(B)\nabla^{d}Z(t)=\theta(B)\nabla^{d}X(t)$, $t=p+d,\ p+d+1,\ldots$. Comparing this with (4.4.2), we see that

$$\nabla^{d}X(t)=M(t), \quad t=p+d-q,\ p+d-q+1,\ldots\ . \tag{4.4.4}$$

So, $\{X(t)\}$ is an integrated white noise. If we have some "observations" on $\{X(t)\}$, then we may set up procedures to test d, as we did in the last section. If we know $\phi(B)$ and $\theta(B)$, then we may obtain such $\{X(t)\}$ from the data, $\{Z(t)\}$. [Notice that only in the case of $\theta(B) = 1$ will this computed $\{X(t)\}$ series correctly replicate the true $\{X(t)\}$. Otherwise, there will generally be errors in the values of $\{X(t)\}$ (especially in the early values which should be abandoned), since the recursion, using (4.4.3), will require unknown initial $X(t)$ to be given assumed values.] But, in practice, none of p, q and the coefficients, ϕ_j and θ_j, are known.

Fortunately, p and q can be identified and all coefficients can be estimated from "observations" on $\{W(t) = \nabla^d Z(t)\}$, and so we are able to get $\hat{\phi}(B)$ and $\hat{\theta}(B)$, estimators of $\phi(B)$ and $\theta(B)$. Suppose

$$\hat{\phi}(\zeta) \neq 0, \quad \hat{\theta}(\zeta) \neq 0, \quad |\zeta| \leq 1, \tag{4.4.5}$$

hold. Then, corresponding to (4.4.3), from data $Z(0), \ldots, Z(n)$, we may obtain $\hat{X}(t)$ by recursion using $\hat{\phi}(B)Z(t) = \hat{\theta}(B)\hat{X}(t)$, $t = 0, 1, \ldots$, setting all the initial values of $Z(t)$ and $\hat{X}(t)$, $t < 0$, zero. This is equivalent to writing

$$\hat{X}(t) = \hat{\theta}(B)^{-1}\hat{\phi}(B)\hat{Z}(t), \quad t = 0, 1, \ldots, n, \tag{4.4.6}$$

where,

$$\hat{Z}(t) = \begin{cases} Z(t), & 0 \leq t \leq n, \\ 0, & t < 0. \end{cases} \tag{4.4.7}$$

From (4.4.6) and (4.4.7), we may put

$$\nabla^d \hat{X}(t) = \hat{M}(t), \quad t = d, d+1, \ldots, n, \tag{4.4.8}$$

where

$$\hat{M}(t) = \hat{\theta}(B)^{-1}\hat{\phi}(B)\hat{W}(t), \quad t \leq n, \tag{4.4.9}$$

with

$$\hat{W}(t) = \begin{cases} W(t), & d \leq t \leq n, \\ \nabla^d \hat{Z}(t), & 0 \leq t < d, \\ 0, & t < 0. \end{cases} \tag{4.4.10}$$

From (4.4.9) and (4.4.10), we see that $\{\hat{M}(t)\}$ is not precisely white noise, so $\{\hat{X}(t)\}$ is not exactly an integrated white noise. In the next subsection we will investigate what happens if we test a hypothesis using the SSPVs of $\{\hat{X}(t)\}$.

A remark should be made. From observations on $\{W(t)\}$, there are many procedures for identifying p and q, such as: Box and Jenkins (1976), Tsay and Tiao (1984), Hannan (1980), Hannan and Rissanen (1982). For a review, see An and Chen (1986). Most of these procedures are consistent. So we may assume that p and q are effectively known (if n is large enough) in our subsequent discussion.

4.4.2 The SSPV from the estimated integrated white noise

Suppose, for a given ARIMA (p, d, q) series under investigation, p, d and q are known. Write $\boldsymbol{\beta}' = (\boldsymbol{\phi}' \quad \boldsymbol{\theta}') = (\phi_1 \cdots \phi_p \; \theta_1 \; \cdots \; \theta_q)$ and denote the true values of the parameters for this particular series by $\dot{\boldsymbol{\beta}}' = (\dot{\boldsymbol{\phi}}' \quad \dot{\boldsymbol{\theta}}') = (\dot{\phi}_1 \cdots \dot{\phi}_p \; \dot{\theta}_1 \cdots \dot{\theta}_q)$. For any $\boldsymbol{\beta}$ with $\phi(\zeta)$ and $\theta(\zeta)$ satisfying (1.1.3), we always have

$$\hat{M}(t, \boldsymbol{\beta}) = \theta(B)^{-1}\phi(B)\hat{W}(t), \quad t = 1, 2, \ldots, n, \tag{4.4.11}$$

where $\hat{W}(t)$ is defined by (4.4.10). Then, $\hat{\boldsymbol{\beta}}$, the least squares estimate (LSE) of $\boldsymbol{\beta}$, minimizes

$$\hat{S}(\boldsymbol{\beta}) = \sum_{t=m+1}^{n} \hat{M}^2(t, \boldsymbol{\beta}). \tag{4.4.12}$$

where $m \geq 0$ is a suitable fixed integer. In (4.4.11) and (4.4.12), we have placed hats on M and S to symbolize that these quantities come from $\{\hat{W}(t)\}$, not from $\{W(t)\}$. Abandoning some early values of $\hat{M}(t, \boldsymbol{\beta})$, $0 \leq t \leq m$, in (4.4.12) has merits in practice, since these values may be widely different from the corresponding values of

$$M(t, \boldsymbol{\beta}) = \theta(B)^{-1}\phi(B)W(t) \tag{4.4.13}$$

[for $t < d$, we do not know $W(t)$]. Then, obviously,

$$M(t, \dot{\boldsymbol{\beta}}) = M(t), \tag{4.4.14}$$

and if $\hat{\boldsymbol{\beta}}$ minimizes (4.4.12), we write

$$\hat{M}(t, \hat{\boldsymbol{\beta}}) = \hat{M}(t). \tag{4.4.15}$$

Following (4.1.5), for a fixed integer $l \geq 0$, write

$$\hat{\gamma}^{(M)}(i) = \frac{1}{n-m} \sum_{t=m+1}^{n} M(t)M(t-i), \quad 0 \leq i \leq l, \tag{4.4.16}$$

$$\hat{\gamma}^{(\hat{M})}(i) = \frac{1}{n-m} \sum_{t=m+1}^{n} \hat{M}(t)\hat{M}(t-i), \quad 0 \le i \le l. \tag{4.4.17}$$

Notice that we can only obtain $\hat{\gamma}^{(\hat{M})}(i)$ given real data, as the SACVF, $\hat{\gamma}^{(M)}(i)$, is unknown in practice, since the $M(t)$ are unknown. In view of (4.5.22) (in the Appendix, below) and then (4.2.51),

$$\hat{\gamma}^{(\hat{M})}(0) = \hat{\gamma}^{(M)}(0) + O(q_n^2) = \sigma^2 + O(q_n) \quad a.s.. \tag{4.4.18}$$

Having got $\hat{\boldsymbol{\beta}}$, the $\hat{X}(t)$ may be obtained from (4.4.6) and satisfy (4.4.8). Then, for $b \in \mathcal{N}$ and $b \ge \max\{d-1, 0\}$, the sample variogram of order $b, \hat{v}_b^{(\hat{X})}(h)$, can be defined in the same way as (4.1.1) and be expressed as [see (4.1.13) and (4.1.14)]

$$\hat{v}_{bd}^{(\hat{X})}(h) = \sum_{r=0}^{h+b-d} \sum_{s=0}^{h+b-d} \pi_{bd}^{h}(r)\pi_{bd}^{h}(s)\hat{\gamma}^{(\hat{M})}(h,r,s), \tag{4.4.19}$$

with

$$\hat{\gamma}^{(\hat{M})}(h,r,s) = \frac{1}{n-h-b+1} \sum_{t=0}^{n-h-b} \hat{M}(h+b+t-r)\hat{M}(h+b+t-s). \tag{4.4.20}$$

By the same discussion as in Lemma 4.2.2 and Theorem 4.2.1, if we can show that, for $r' = h+b-r, \ s' = h+b-s, \ (h,r,s) \in \mathcal{D}$,

$$\hat{M}(t+r')\hat{M}(t+s') = o(n^{1/2}) \quad a.s., \tag{4.4.21}$$

then

$$\hat{v}_{bd}^{(\hat{X})}(h) = \sum_{i=0}^{h+b-d} g_{bd}(h,i)\hat{\gamma}^{(\hat{M})}(i) + o(n^{1/2}) \quad a.s.. \tag{4.4.22}$$

In fact, (4.4.21) is true since $M(t+r')M(t+s') = o(n^{1/2})$ *a.s.*, and

$$\hat{M}(t+r')\hat{M}(t+s') - M(t+r')M(t+s') = o(n^{1/4}) \quad a.s., \tag{4.4.23}$$

as is proved in the Appendix (see subsection 5.5.2). So (4.2.22) holds.

Let $\hat{\rho}_b^{(\hat{X})}(h)$ and $\hat{\rho}_b^{*(\hat{X})}(h)$ be the SSPVs of $\{\hat{X}(t)\}$ which are defined from $\hat{v}_b^{(\hat{X})}(h)$ and $\hat{\gamma}^{(\hat{M})}(0)$ [see (4.1.2), (4.1.4), (4.1.8) and (4.1.10)]. Similarly to the way that (4.2.21) and (4.2.40) was derived from (4.2.6), and noticing (4.4.18), (4.4.22) leads to

$$\hat{\boldsymbol{\rho}}_{bd}^{(\hat{X})} = G_{bd}\hat{\boldsymbol{\rho}}^{(\hat{M})} + o(n^{-1/2}), \quad \hat{\boldsymbol{\rho}}_{bd}^{*(\hat{X})} = (-1)^{b+1}G_{bd}^{*}\hat{\boldsymbol{\rho}}^{(\hat{M})} + o(n^{-1/2}), \ a.s.; \tag{4.4.24}$$

where

$$\tilde{\rho}_{bd}^{(\hat{X})} = \hat{\rho}_{bd}^{(\hat{X})} - \rho_{bd}^{(X)}, \quad \tilde{\rho}_{bd}^{*(\hat{X})} = \hat{\rho}_{bd}^{*(\hat{X})} - \rho_{bd}^{*(X)}, \tag{4.4.25}$$

$$\tilde{\rho}^{(\hat{M})} = \hat{\rho}^{(\hat{M})} - \mathbf{0} = (\hat{\gamma}^{(\hat{M})}(1) \ \cdots \ \hat{\gamma}^{(\hat{M})}(k+b-d))'/\hat{\gamma}^{(\hat{M})}(0). \tag{4.4.26}$$

The elements in $\rho_{bd}^{(X)}$ and $\rho_{bd}^{*(X)}$ are given by (4.3.1) since the $X(t)$, which are estimated by the $\hat{X}(t)$, are integrated white noise [see (4.4.4)]. So $\tilde{\rho}_{bd}^{(\hat{X})}$ and $\tilde{\rho}_{bd}^{*(\hat{X})}$ may be obtained from data and used as test statistics, provided that we can derive their asymptotic properties. In the following, for simplicity of statement, we only give results for $\tilde{\rho}_{bd}^{(\hat{X})}$. The results for $\tilde{\rho}_{bd}^{*(\hat{X})}$ are the same except that, instead of G_{bd}, the transformation matrix $(-1)^{b+1}G_{bd}^{*}$ should be used [see (4.2.40)].

In view of (4.4.24), it is sufficient to establish the asymptotic properties for $\tilde{\rho}^{(\hat{M})}$, the SACF for the residuals of the estimated model. Write

$$\hat{\gamma}^{(\hat{M})} = (\hat{\gamma}^{(\hat{M})}(1) \quad \cdots \quad \hat{\gamma}^{(\hat{M})}(k+b-d))'. \tag{4.4.27}$$

Then, from (4.4.18), (4.4.26) and then (4.4.24), we get

$$\tilde{\rho}_{bd}^{(\hat{X})} = \sigma^{-2} G_{bd} \hat{\gamma}^{(\hat{M})} \{1 + O(q_n)\} + o(n^{-1/2}) \quad a.s.\,. \tag{4.4.28}$$

Concerning the asymptotic properties of $\hat{\gamma}^{(\hat{M})}$, the main difficulty is that the errors in the coefficient estimates are involved. However, for an ARMA series (1.1.2), when the stochastic shocks $A(t)$ are i.i.d., McLeod (1978) has already derived the asymptotic distribution of $\hat{\gamma}^{(A)}$ [defined for the $\hat{A}(t)$ as $\hat{\gamma}^{(\hat{M})}$ was for the $\hat{M}(t)$]. Since we now need to extend the i.i.d. case to that of martingale differences and, moreover, want to obtain the rate of almost sure convergence, we require a result which is stronger and holds under more general conditions than that of McLeod (1978).

Consider an ARMA series (4.4.1), where $\phi(B)$ and $\theta(B)$ are given by (1.1.3). Again denote the true values of the parameters for a given series, $\{W(t)\}$, by $\dot{\beta}' = (\dot{\phi}' \ \dot{\theta}')$, and the corresponding polynomial operators in B by $\dot{\phi}(B)$ and $\dot{\theta}(B)$. Denote the Taylor expansions of $\phi^{-1}(B)$ and $\theta^{-1}(B)$ by

$$\phi^{-1}(B) = \sum_{s=0}^{\infty} \phi_s^{*} B^{s}, \quad \theta^{-1}(B) = \sum_{s=0}^{\infty} \theta_s^{*} B^{s}. \tag{4.4.29}$$

So the coefficients in $\dot{\phi}^{-1}(B)$ and $\dot{\theta}^{-1}(B)$ should be denoted by $\dot{\phi}_s^*$ and $\dot{\theta}_s^*$, respectively. Let $l = k + b - d$, and define:

$$\Psi = \begin{pmatrix} -\dot{\phi}_0^* & \cdots & -\dot{\phi}_{1-p}^* & \dot{\theta}_0^* & \cdots & \dot{\theta}_{1-q}^* \\ \vdots & & \vdots & \vdots & & \vdots \\ -\dot{\phi}_{l-1}^* & \cdots & -\dot{\phi}_{l-p}^* & \dot{\theta}_{l-1}^* & \cdots & \dot{\theta}_{l-q}^* \end{pmatrix}_{l\times(p+q)} \tag{4.4.30}$$

$$J = \begin{pmatrix} (\sum_{s=0}^{\infty} \dot{\phi}_{s-i}^* \dot{\phi}_{s-j}^*)_{p\times p} & (-\sum_{s=0}^{\infty} \dot{\phi}_{s-i}^* \dot{\theta}_{s-j}^*)_{p\times q} \\ (-\sum_{s=0}^{\infty} \dot{\theta}_{s-i}^* \dot{\phi}_{s-j}^*)_{q\times p} & (\sum_{s=0}^{\infty} \dot{\theta}_{s-i}^* \dot{\theta}_{s-j}^*)_{q\times q} \end{pmatrix}, \tag{4.4.31}$$

where the dimension and the (i,j)-th element of each submatrix as indicated. In both definitons, $\dot{\phi}_s^* = 0$, $\dot{\theta}_s^* = 0$, if $s < 0$.

Theorem 4.4.1 *Suppose the $W(t)$ satisfy (4.4.1) and the true values of the parameters are $\dot{\boldsymbol{\beta}}$, where the $M(t)$ are strictly stationary, ergodic martingale differences given by (4.2.48) and satisfying (4.2.49). Suppose $\hat{\boldsymbol{\beta}}$ minimizes (4.4.12) and $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \dot{\boldsymbol{\beta}}$. Then, for any constant vector $\boldsymbol{\lambda}$ of dimension $p + q$,*

$$\limsup_{n\to\infty} \{n/(2\boldsymbol{\lambda}' J^{-1} \boldsymbol{\lambda} \log\log n)\}^{1/2} \mid \boldsymbol{\lambda}' \tilde{\boldsymbol{\beta}} \mid = 1 \quad a.s. \quad . \tag{4.4.32}$$

Moreover, using this $\hat{\boldsymbol{\beta}}$, suppose $\hat{\gamma}^{(\dot{M})}$ is defined by (4.4.27) and (4.4.17), $l = k + b - d$, and let

$$\mathbf{N}(t) = \begin{pmatrix} M(t-1) \\ \vdots \\ M(t-l) \end{pmatrix} - \Psi J^{-1} \begin{pmatrix} -\dot{\phi}^{-1}(B)(M(t-1) & \cdots & M(t-p))' \\ \dot{\theta}^{-1}(B)(M(t-1) & \cdots & M(t-q))' \end{pmatrix}, \tag{4.4.33}$$

and

$$\mathbf{L}(t) = \mathbf{N}(t) M(t). \tag{4.4.34}$$

Then

$$\hat{\gamma}^{(\dot{M})} = \frac{1}{n-m} \{ \sum_{t=m+1}^{n} \mathbf{L}(t) \} + O(q_n^2) \quad a.s., \tag{4.4.35}$$

$$E\{\mathbf{L}(s)\mathbf{L}'(t)\} = \sigma^4 (I_l - \Psi J^{-1} \Psi') \delta_{s-t}, \tag{4.4.36}$$

where $q_n = (n^{-1} \log\log n)^{1/2}, \delta_s$ is the Kronecker delta function and I_l is the unit matrix of order l.

Under the conditions of this theorem, the CLT for $\hat{\boldsymbol{\beta}}$ is well-known (Hannan 1973), but the LIL for $\hat{\boldsymbol{\beta}}$, (4.4.32), has not been given previously. The theorem is proved in the Appendix (subsection 4.5.1). The main clue for deriving (4.4.35) comes from McLeod (1978). Considering the SSPV of the estimated integrated white noise, $\{\hat{X}(t)\}$, we have the following theorem.

Theorem 4.4.2 *Suppose $d \in \mathcal{N}$ and $\{Z(t)\} \in \tilde{I}_d$, with its hub series $\{W(t)\}$ and parameter estimates $\hat{\boldsymbol{\beta}}$ described in Theorem 4.4.1, and the $\hat{X}(t)$ are defined by (4.4.6). Instead of (4.3.10), write*

$$C_{bd} = G_{bd}(I_l - \Psi J^{-1}\Psi')G'_{bd}. \tag{4.4.37}$$

Then [for the definition, see (4.4.25)]:

$$n^{1/2}\tilde{\boldsymbol{\rho}}_{bd}^{(\hat{X})} \xrightarrow{L} N(\mathbf{0}, C_{bd}) \tag{4.4.38}$$

$$\limsup_{n\to\infty}\{n/(2\log\log n)\}^{1/2} \mid \tilde{\rho}_{bd}^{(\hat{X})}(h) \mid = c_{bd}(h,h) \ a.s., \ 1+\delta \le h \le k, \tag{4.4.39}$$

where δ is defined by (4.3.12) and $c_{bd}(h,h)$ is the $(h-\delta)$-th diagonal element of $G_{bd}(I_l - \Psi J^{-1}\Psi')G'_{bd}$. [Again, when n is not very large, using an $n^{1/2}\tilde{\tilde{\boldsymbol{\rho}}}_{bd}^{(\hat{X})}$ analogous to (4.2.25) instead of $n^{1/2}\tilde{\boldsymbol{\rho}}_{bd}^{(\hat{X})}$, and $n-h$ rather than n in (4.4.39), may both give better approximations and so be preferable.]

Proof Given any constant vector $\boldsymbol{\mu}$ of dimension l, the $\boldsymbol{\mu}'\mathbf{L}(t)$ are strictly stationary, ergodic and square-integrable martingale differences, and then the CLT and the LIL of the $\boldsymbol{\mu}'\mathbf{L}(t)$ follow similarly as in the proof of Lemma 4.3.1; and, in view of (4.4.36), the asymptotic variance of $\boldsymbol{\mu}'\mathbf{L}(t)$ is $\sigma^4\boldsymbol{\mu}'(I_l - \Psi J^{-1}\Psi')\boldsymbol{\mu}$.

Let $\boldsymbol{\xi}'$ be any constant vector of dimension $k-\delta$. Then, from (4.4.28) and (4.4.35),

$$\boldsymbol{\xi}'\tilde{\boldsymbol{\rho}}_{bd}^{(\hat{X})} = [\{\frac{\sigma^{-2}}{n-m}\sum_{t=m+1}^{n}\boldsymbol{\xi}'G_{bd}\mathbf{L}(t)\} + O(q_n^2)]\{1+O(q_n)\} + o(n^{-1/2}) \ a.s. \tag{4.4.40}$$

So (4.4.38) follows from (4.4.40), the CLT for $\boldsymbol{\mu}'\mathbf{L}(t) = \boldsymbol{\xi}'G_{bd}\mathbf{L}(t)$ and using the Cramér-Wold device. Let $\boldsymbol{\xi}$ be the null vector with $(h-\delta)$-th element replaced by 1. Then (4.4.39) follows from (4.4.40) and the LIL for $\boldsymbol{\mu}'\mathbf{L}(t) = \boldsymbol{\xi}'G_{bd}\mathbf{L}(t)$. □

4.4.3 The errors in the parameter estimates

The procedures, using (4.4.38) and (4.4.39) (or the corresponding results for $\bar{\rho}_{bd}^{(\hat{X})}$) to test $H_0^{(d)}: \{Z(t)\} \in \tilde{I}_d$, are almost the same as in subsections 4.3.3. and 4.3.4, with C_{bd} defined by (4.4.37) instead of (4.3.10). The only problem is that the elements of Ψ and J [see (4.4.30) and (4.4.31)] are unknown. However, the next theorem shows that the errors of the estimates for these unknown elements are $O(q_n)$ *a.s.* . Let $\hat{\Psi}$ and $\hat{J}$ be formed from Ψ and J by replacing the parameters by their estimates. Then, since $\mid J \mid > 0$ [as J is a variance-covariance matrix of $p+q$ linearly independent random variables — see (4.5.8)], it is easy to show:

$$\hat{C}_{bd} \stackrel{def}{=} G_{bd}(I_l - \hat{\Psi}\hat{J}^{-1}\hat{\Psi}')G'_{bd} = C_{bd} + O(q_n) \quad a.s., \tag{4.4.41}$$

where C_{bd} is as defined in (4.4.37). So, for test statistics such as (4.3.24) or (4.3.26), when C_{bd} is replaced by $\hat{C}_{bd}$, the asymptotic properties given by (4.4.38) and (4.4.39) do not change, and hence the test procedures may be carried out as before.

Theorem 4.4.3 *Under the same assumptions as in Theorem 4.4.1, and with $\hat{\phi}^*_s, \hat{\theta}^*_s, \dot{\phi}^*_s$ and $\dot{\theta}^*_i$ as defined by (4.4.29) on replacing ϕ and θ, respectively, by $\hat{\phi}$ and $\hat{\theta}$ and by $\dot{\phi}$ and $\dot{\theta}$, and denoting the coefficients of $\hat{\phi}(B)$ and $\hat{\theta}(B)$ by $\hat{\beta}' = (\hat{\boldsymbol{\phi}}' \ \hat{\boldsymbol{\theta}}')$:*

$$\hat{\phi}^*_i - \dot{\phi}^*_i = O(q_n) \quad a.s., \tag{4.4.42}$$

$$\sum_{s=0}^{\infty} \hat{\phi}^*_{s-i}\hat{\phi}^*_{s-j} - \sum_{s=0}^{\infty} \dot{\phi}^*_{s-i}\dot{\phi}^*_{s-j} = O(q_n) \quad a.s., \tag{4.4.43}$$

with corresponding results also holding when ϕ is replaced everywhere by θ, and when just the second ϕ of every product in (4.4.43) is replaced by θ.

Proof In view of (4.4.32),

$$\hat{\boldsymbol{\beta}} - \dot{\boldsymbol{\beta}} = O(q_n) \quad a.s. . \tag{4.4.44}$$

So, for sufficiently large n, $\hat{\boldsymbol{\beta}}$ falls in a closed neighbourhood of $\dot{\boldsymbol{\beta}}$ throughout which condition (1.1.3) holds and

$$\mid \phi^*_s \mid \leq c_1 e^{-c_2 s} \tag{4.4.45}$$

for some constants $c_1 > 0,\ c_2 > 0$. Next consider

$$\hat{\phi}^{-1}(B) - \dot{\phi}^{-1}(B) = \hat{\phi}^{-1}(B)\dot{\phi}^{-1}(B)\{\dot{\phi}(B) - \hat{\phi}(B)\}, \tag{4.4.46}$$

and write $\hat{\phi}^{-1}(B)\dot{\phi}^{-1}(B) = \sum_{s=0}^{\infty} \hat{\psi}_s B^s$. Then, for sufficiently large n,

$$|\hat{\psi}_s| \leq \sum_{i+j=s} |\hat{\phi}_i^* \dot{\phi}_j^*| \leq c_1^2(s+1)e^{-c_2 s} \leq c_3 e^{-c_4 s} \quad a.s., \tag{4.4.47}$$

for constants $c_3 > 0,\ c_4 > 0$ ($c_4 < c_2$). So, from (4.4.44), (4.4.46) and (4.4.47),

$$|\hat{\phi}_s^* - \dot{\phi}_s^*| \leq e^{-c_4 s} O(q_n) \quad a.s., \tag{4.4.48}$$

and (4.4.42) follows immediately. Rewrite the LHS of (4.4.43) as

$$\sum_{s=0}^{\infty} \hat{\phi}_{s-i}^*(\hat{\phi}_{s-j}^* - \dot{\phi}_{s-j}^*) + \sum_{s=0}^{\infty}(\hat{\phi}_{s-i}^* - \dot{\phi}_{s-i}^*)\dot{\phi}_{s-j}^*,$$

then (4.4.43) follows from (4.4.45) and (4.4.48). □

4.4.4 The test procedure

Although the described technique can transform an ARIMA series to an approximate integrated white noise series, this technique can not detect underdifferencing. For example, suppose $\{Z(t)\}$ satisfies the model

$$(1 - \phi_1 B)Z(t) = M(t), \quad t = 1, 2, \ldots, \tag{4.4.49}$$

and the true value of ϕ_1 is 1 [i.e. the true model is $\nabla Z(t) = M(t)$]. Allow us to set up the (mistaken) null hypothesis of $d = 0$ [i.e. model (4.4.49) with $|\phi_1| < 1$] and, based on this model, carry out the transformation procedure. Then (4.4.6) becomes $\hat{X}(t) = (1 - \hat{\phi}_1 B)\hat{Z}(t)$. Notice that, since in fact $\phi_1 = 1, \hat{\phi}_1 - 1 = O_p(1/n)$ (see Dickey and Fuller 1979), and the estimate of ϕ_1, when $\phi_1 = 1$, is much more precise than when $|\phi_1| < 1$). So the $\hat{X}(t) = \hat{M}(t)$ are very close to the $M(t)$ (when $t > 0$), and the null hypothesis of $d = 0$ (which was originally false) is now approximately true for the transformed model, and so is very likely to be accepted by the test procedure.

To avoid such a situation, we suggest the following modification which ensures that the hypothesis tests are not applied to underdifferenced series.

Suppose the true value of d is d'. Guess a safe upper bound for d', say $d_0 \geq d'$. Let $b = d_0$, then $\gamma_{bd'}(h)$ (for sake of argument) has a horizontal line as its asymptote, and hence $\hat{\gamma}_b(h)$ should also appear to have such an asymptote. Now iteratively decrement b by one until, for some (first) $b \in \mathcal{N}, \hat{\gamma}_b(h)$ no longer appears to have a horizontal line for its asymptote or at least it is ambiguous as to whether the slope is still zero. Then let $d = b + 1$ and set up the hypothesis $H_0^{(d)} : \{Z(t)\} \in \tilde{I}_d$. By doing it this way, we are dealing with either $d = d'$ or $d > d'$ (the possibility of $d < d'$ is excluded).

When $d = d'$, it is appropriate to test $H_0^{(d)}$ with the procedures proposed in the last section [but based on Theorem 4.4.2 and (4.4.41)].

Let us see what could happen when $d > d'$. For sake of argument, let $d = d' + 1$. In this situation, (4.4.2) has the form

$$\phi(B)\nabla^{d'+1}Z(t) = \theta(B)M(t). \tag{4.4.50}$$

Since d' is the true degree, $\theta(B)$ must have a factor of $\nabla = 1 - B$. Write $\nabla^{d'+1}Z(t) = W_*(t)$, where $\{W_*(t) = \nabla W(t)\}$ is not the hub series $\{W(t)\}$. Using values of $W_*(t)$, obtained from the observed $Z(t)$, we can estimate $\beta' = (\phi' \; \theta')$ in (4.4.50) by $\hat{\beta}' = (\hat{\phi}' \; \hat{\theta}')$. Then the $\hat{M}(t)$ may be obtained from the following difference equation [cf (4.4.9)]

$$\hat{\theta}(B)\hat{M}(t) = \hat{\phi}(B)\hat{W}_*(t) \tag{4.4.51}$$

with $\hat{W}_*(t)$ defined analogously to (4.4.10). Put (4.4.50) in the form

$$\theta(B)M(t) = \phi(B)W_*(t). \tag{4.4.52}$$

If $\hat{\beta}$ is a poor estimate of β, comparing (4.4.51) with (4.4.52), we see that the $\hat{M}(t)$ are usually far from the $M(t)$. If $\hat{\beta}$ is a good estimate of β, then $\hat{\theta}(\zeta)$ has a zero near to 1, and hence (4.4.51) is almost an unstable difference equation for obtaining $\hat{M}(t)$ (by recursion). Then the influence of any perturbation in a $\hat{\phi}(B)\hat{W}_*(t)$, at an early t, on the $\hat{M}(t)$ of later t persists, and again the $\hat{M}(t)$ will usually be far from the $M(t)$. So, testing whether the $\hat{X}(t)$ [given by (4.4.6)] are approximately d-times integrated white noise [see (4.4.8)] will usually result in d $(= d' + 1)$ being rejected.

In general, should $H_0^{(d)}$ be rejected, then we consider $H_0^{(d-1)}$, and so on down to the first non-negative integer, say $\hat{d}$, such that $H_0^{(\hat{d})}$ is accepted by the testing. Then $\hat{d}$ is the estimate of d'.

4.5 Appendix: Proofs of some Section 4.4 Results

4.5.1 The proof of Theorem 4.4.1

Lemma 4.5.1 *Suppose the $M(t)$ are defined by the material (4.2.48) through (4.2.49).*

$$Y_i(t) = \sum_{s=0}^{\infty} \alpha_{is} M(t-s), \quad i = 1, 2; \quad t = \ldots, -1, 0, 1, \ldots;$$

and, as $s \to \infty$, $|\alpha_{is}| \to 0$ geometrically. Then

$$\frac{1}{n}\sum_{t=1}^{n} Y_1(t)Y_2(t) = E\{Y_1(t)Y_2(t)\} + O(q_n) \quad a.s. \quad .$$

This lemma can be regarded as a corollary of Theorem 2 in An et al (1982).

Corresponding to (4.4.12), write

$$S(\boldsymbol{\beta}) = \sum_{t=m+1}^{n} M^2(t, \boldsymbol{\beta}), \tag{4.5.1}$$

where $m \in \mathcal{N}$ is fixed and the $M(t, \boldsymbol{\beta})$, defined by (4.4.13), are unknown in practice. Then $S(\boldsymbol{\beta})$ is also unknown, but we shall see that it can be closely approximated by $\hat{S}(\boldsymbol{\beta})$.

In the following, for simplicity of notation, we always replace any $n - m$ divisor by n (which makes no difference asymptotically).

Consider the minimization of (4.5.1) in a closed neighbourhood $\mathcal{B}$ of $\dot{\boldsymbol{\beta}}$, throughout which condition (1.1.3) holds uniformly [i.e., $\phi(B)$ and $\theta(B)$ are bounded away from zero]. For $\boldsymbol{\beta} \in \mathcal{B}$, define $U(t, \boldsymbol{\beta})$ and $V(t, \boldsymbol{\beta})$ as follows:

$$M(t, \boldsymbol{\beta}) = -\phi(B)U(t, \boldsymbol{\beta}), \tag{4.5.2a}$$

$$M(t, \boldsymbol{\beta}) = \theta(B)V(t, \boldsymbol{\beta}). \tag{4.5.2b}$$

Differentiating (4.4.13), with reference to (1.1.3), and then using (4.4.13) followed by (4.5.2), we get:

$$\partial M(t,\boldsymbol{\beta})/\partial\phi_i = -\theta^{-1}(B)W(t-i) = U(t-i,\boldsymbol{\beta}) \tag{4.5.3a}$$

$$\partial M(t,\boldsymbol{\beta})/\partial\theta_i = \theta^{-2}(B)\phi(B)W(t-i) = V(t-i,\boldsymbol{\beta}). \tag{4.5.3b}$$

Then, differentiating (4.5.1) yields:

$$\partial S(\boldsymbol{\beta})/\partial\phi_i = 2\sum_{t=m+1}^{n} U(t-i,\boldsymbol{\beta})M(t,\boldsymbol{\beta}) \tag{4.5.4a}$$

$$\partial S(\boldsymbol{\beta})/\partial\theta_i = 2\sum_{t=m+1}^{n} V(t-i,\boldsymbol{\beta})M(t,\boldsymbol{\beta}); \tag{4.5.4b}$$

and we will use a notation exemplified by writing the value of $\partial S(\boldsymbol{\beta})/\partial\phi_i$, at $\boldsymbol{\beta} = \dot{\boldsymbol{\beta}}$, as $\partial S(\dot{\boldsymbol{\beta}})/\partial\phi_i$. Since $M(t,\dot{\boldsymbol{\beta}}) \equiv M(t)$, (4.5.2) gives $E\{U(t-i,\dot{\boldsymbol{\beta}})M(t,\dot{\boldsymbol{\beta}})\} = 0$ and $E\{V(t-i,\dot{\boldsymbol{\beta}})M(t,\dot{\boldsymbol{\beta}})\} = 0$. So, using Lemma 4.5.1, we get

$$n^{-1}\partial S(\dot{\boldsymbol{\beta}})/\partial\boldsymbol{\beta} = O(q_n) \quad a.s. \quad . \tag{4.5.5}$$

Next, differentiating (4.5.4) and (4.5.2) and using (4.5.3) gives

$$\partial^2 S(\dot{\boldsymbol{\beta}})/(\partial\phi_i\partial\phi_j) = 2\sum_{t=m+1}^{n} U(t-i,\dot{\boldsymbol{\beta}})U(t-j,\dot{\boldsymbol{\beta}}) \tag{4.5.6a}$$

$$n^{-1}\partial^2 S(\dot{\boldsymbol{\beta}})/(\partial\phi_i\partial\theta_j) = 2n^{-1}\sum_{t=m+1}^{n} U(t-i,\dot{\boldsymbol{\beta}})V(t-j,\dot{\boldsymbol{\beta}}) + O(q_n) \quad a.s. \tag{4.5.6b}$$

$$n^{-1}\partial^2 S(\dot{\boldsymbol{\beta}})/(\partial\theta_i\partial\theta_j) = 2n^{-1}\sum_{t=m+1}^{n} V(t-i,\dot{\boldsymbol{\beta}})V(t-j,\dot{\boldsymbol{\beta}}) + O(q_n) \quad a.s., \tag{4.5.6c}$$

where the remainders of $O(q_n)$ $a.s.$ follow from Lemma 4.5.1. For instance, the remainder in (4.5.6b) is [on, say, differentiating (4.5.4a) and using (4.5.3b), and then differentiating (4.5.2a) and using (4.5.3b) again]:

$$\frac{2}{n}\sum_{t=m+1}^{n} \frac{\partial U(t-i,\dot{\boldsymbol{\beta}})}{\partial\theta_j}M(t) = -\frac{2}{n}\sum_{t=m+1}^{n} \{\theta^{-2}(B)W(t-i-j)\}M(t) = O(q_n) \quad a.s..$$

When $\boldsymbol{\beta} = \dot{\boldsymbol{\beta}}$, (4.5.2) becomes

$$U(t,\dot{\boldsymbol{\beta}}) = -\dot{\phi}^{-1}(B)M(t), \quad V(t,\dot{\boldsymbol{\beta}}) = \dot{\theta}^{-1}(B)M(t), \tag{4.5.7}$$

and then

$$\mathrm{Var}(U(t-1,\dot{\boldsymbol{\beta}}) \ \cdots \ U(t-p,\dot{\boldsymbol{\beta}}) \quad V(t-1,\dot{\boldsymbol{\beta}}) \ \cdots \ V(t-q,\dot{\boldsymbol{\beta}}))' = \sigma^2 J. \tag{4.5.8}$$

Using Lemma 4.5.1, (4.5.6), (4.5.7) and (4.5.8) then give

$$n^{-1}\partial^2 S(\dot{\beta})/(\partial\beta\partial\beta') = 2\sigma^2 J + O(q_n) \quad a.s., \tag{4.5.9}$$

where the LHS denotes a matrix having (i,j)-th element $n^{-1}\partial^2 S(\dot{\beta})/(\partial\beta_i\partial\beta_j)$ ($\beta_i = \phi_i, 1 \le i \le p;\ \beta_{p+i} = \theta_i, 1 \le i \le q$).

Considering the third order derivatives of $S(\beta)/n$ with respect to β_i, β_j and β_k, in $\mathcal{B}$, we see that $\partial^3\{M^2(t,\beta)\}/(\partial\beta_i\partial\beta_j\partial\beta_k)$ can be expressed as a finite sum of products of two factors, each having the form

$$a\theta^{-g}(B)\phi^h(B)W(t-l) = a\theta^{-g}(B)\phi^h(B)\dot{\phi}^{-1}(B)\dot{\theta}(B)M(t-l),$$

where a, g, h, l are integers, $g \ge 0, h \ge 0, l > 0$. So each factor can be expressed as $Y_i(t-l)$ in Lemma 4.5.1 which gives

$$\frac{1}{n}\frac{\partial^3 S(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_k} = E\{\frac{\partial^3 M^2(t,\beta)}{\partial\beta_i\partial\beta_j\partial\beta_k}\} + O(q_n) \quad a.s., \tag{4.5.10}$$

where the leading term on the RHS is bounded in $\mathcal{B}$.

Now consider the following equation

$$n^{-1}\partial S(\beta)/\partial\beta = 0 \tag{4.5.11}$$

for $\beta \in \mathcal{B}$. Using a Taylor expansion, (4.5.11) becomes

$$\frac{1}{n}\frac{\partial S(\dot{\beta})}{\partial\beta} + \frac{1}{n}\frac{\partial^2 S(\dot{\beta})}{\partial\beta\partial\beta'}(\beta - \dot{\beta})$$

$$+\frac{1}{2n}((\beta-\dot{\beta})'\frac{\partial^3 S(\beta^{(1)})}{\partial\beta_1\partial\beta\partial\beta'}(\beta-\dot{\beta})\cdots(\beta-\dot{\beta})'\frac{\partial^3 S(\beta^{(p+q)})}{\partial\beta_{p+q}\partial\beta\partial\beta'}(\beta-\dot{\beta}))' = 0, \tag{4.5.12}$$

where all the $\beta^{(i)}$ lie within the join of β and $\dot{\beta}$. In view of (4.5.5), (4.5.9), and the boundedness of (4.5.10), carrying through a typical proof of the existence of the maximum likelihood estimates (MLE), we know that (4.5.11) has a solution $\hat{\beta}$ satisfying $\hat{\beta} - \dot{\beta} = O(q_n)$ $a.s.$ and which can in fact be expressed as

$$\hat{\beta} - \dot{\beta} = -(2\sigma^2 J)^{-1} n^{-1}\partial S(\dot{\beta})/\partial\beta + O(q_n^2)$$

$$= -\sigma^2 J^{-1}\frac{1}{n}\sum_{t=m+1}^{n}\begin{pmatrix} -\dot{\phi}^{-1}(B)(M(t-i) & \cdots & M(t-p))' \\ \dot{\theta}^{-1}(B)(M(t-i) & \cdots & M(t-q))' \end{pmatrix} M(t) + O(q_n^2) \quad a.s.\ , \tag{4.5.13}$$

on using (4.5.4) followed by (4.5.2).

Next, we show that, (4.5.13) is still true if $\hat{\boldsymbol{\beta}}$ is obtained from

$$n^{-1}\partial\hat{S}(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0} \qquad (\boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}), \tag{4.5.14}$$

instead of (4.5.11), as $S(\boldsymbol{\beta})$ is not known in practice.

Similarly to (4.5.2), we may define $\hat{U}(t,\boldsymbol{\beta})$ and $\hat{V}(t,\boldsymbol{\beta})$ from $\hat{M}(t,\boldsymbol{\beta})$ [see (4.4.11)], and analogously to (4.4.10) write

$$\tilde{W}(t) = W(t) - \hat{W}(t) = \begin{cases} 0, & t \geq d, \\ W(t) - \nabla^d \hat{Z}(t), & 0 \leq t < d, \\ W(t), & t < 0. \end{cases} \tag{4.5.15}$$

Consider, say,

$$\frac{\partial\hat{S}(\boldsymbol{\beta})}{\partial\phi_i} - \frac{\partial S(\boldsymbol{\beta})}{\partial\phi_i} = 2\sum_{t=m+1}^{n}\{\hat{U}(t-i,\boldsymbol{\beta})\hat{M}(t,\boldsymbol{\beta}) - U(t-i,\boldsymbol{\beta})M(t,\boldsymbol{\beta})\}$$

$$= 2\sum_{t=m+1}^{n}[\hat{U}(t-i,\boldsymbol{\beta})\{\hat{M}(t,\boldsymbol{\beta}) - M(t,\boldsymbol{\beta})\} + \{\hat{U}(t-i,\boldsymbol{\beta}) - U(t-i,\boldsymbol{\beta})\}M(t,\boldsymbol{\beta})]$$

$$= 2\sum_{t=m+1}^{n}[\{-\theta^{-1}(B)\hat{W}(t)\}\{-\theta^{-1}(B)\phi(B)\tilde{W}(t)\} - \{\theta^{-1}(B)\tilde{W}(t)\}M(t,\boldsymbol{\beta})]. \tag{4.5.16}$$

For $\boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}$, again write $\theta^{-1}(B) = \sum_{s=0}^{\infty}\theta_s^{*}B^{s}$ and write $\theta^{-1}(B)\phi(B) = \sum_{s=0}^{\infty}\zeta_s B^{s}$. Then θ_s^{*} and ζ_s are dominated similarly to the way that ϕ_s^{*} is [see (4.4.45)]. By Lemma 4.2.1, due to $E\{W^4(t)\} < \infty$ and the stationarity, $W(t) = o(|\, t\,|^{1/4})$ *a.s.*, so:

$$|\,\theta^{-1}(B)\hat{W}(t)\,| = |\sum_{s=0}^{t-d}\theta_s^{*}W(t-s)\,| + |\sum_{s=0}^{d-1}\theta_{t-s}^{*}\nabla^d\hat{Z}(s)\,| = o(t^{1/4}) \quad a.s.$$

$$|\,\theta^{-1}(B)\phi(B)\tilde{W}(t)\,| \leq |\sum_{s=0}^{d-1}\zeta_{t-s}\nabla^d\hat{Z}(s)\,| + |\sum_{s=t+1}^{\infty}\zeta_s W(t-s)\,|$$

$$\leq O(e^{-c_1 t}) + o(\sum_{s=1}^{\infty}e^{-c_1(t+s)}s^{1/4}) = O(e^{-c_1 t}) \quad a.s. \quad ,$$

where this last result can be specialised for the case $\phi(B) = 1$. So, the RHS of (4.5.16) is of order $\sum_{t=m+1}^{n}o(e^{-c_1 t}t^{1/4}) = O(1)$ *a.s.*. That is

$$\partial\hat{S}(\boldsymbol{\beta})/\partial\phi_i - \partial S(\boldsymbol{\beta})/\partial\phi_i = O(1) \quad a.s.\,. \tag{4.5.17}$$

In a similar manner, we can show that corresponding results hold when ϕ is replaced by θ in this formula, and that the differences between all corresponding second (and third) derivatives of $\hat{S}(\boldsymbol{\beta})$ and $S(\boldsymbol{\beta})$ are also $O(1)$. Now, analogously to (4.5.12), we may write (4.5.14) as

$$\frac{1}{n}\frac{\partial \hat{S}(\dot{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} + \frac{1}{n}\frac{\partial^2 \hat{S}(\dot{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}(\boldsymbol{\beta} - \dot{\boldsymbol{\beta}})$$

$$+\frac{1}{2n}((\boldsymbol{\beta}-\dot{\boldsymbol{\beta}})'(\frac{\partial^3 \hat{S}(\boldsymbol{\beta}^{(1)})}{\partial \beta_1 \partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'})(\boldsymbol{\beta}-\dot{\boldsymbol{\beta}})\cdots(\boldsymbol{\beta}-\dot{\boldsymbol{\beta}})'(\frac{\partial^3 \hat{S}(\boldsymbol{\beta}^{(p+q)})}{\partial \beta_{p+q} \partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'})(\boldsymbol{\beta}-\dot{\boldsymbol{\beta}}))' = \mathbf{0}. \quad (4.5.18)$$

Then considering (4.5.18), if $\hat{S}$ is replaced by S in the LHS, the RHS becomes "$O(1/n)$ *a.s.*". That means a solution, $\hat{\boldsymbol{\beta}}$, of (4.5.14) satisfies (4.5.12) if the RHS of (4.5.12) is modified to $O(1/n)$ *a.s.* . However, this modification does not affect the derivation of (4.5.13); so (4.5.13) also holds for the solution of (4.5.14).

Given any constant vector $\boldsymbol{\mu}$ of dimension $p+q$, the

$$\boldsymbol{\mu}' \begin{pmatrix} -\dot{\phi}^{-1}(B)(M(t-1) & \cdots & M(t-p))' \\ \dot{\theta}^{-1}(B)(M(t-1) & \cdots & M(t-q))' \end{pmatrix} M(t) \qquad (4.5.19)$$

are strictly stationary, ergodic and square-integrable martingale differences, so the LIL holds (Stout, 1970). Then (4.4.32) follows from (4.5.13), on putting $\boldsymbol{\mu} = J^{-1}\boldsymbol{\lambda}$.

In the following, the material from (4.4.11) to (4.4.17) inclusive is used repeatedly.

$$\hat{\gamma}^{(\hat{M})}(i) - \frac{1}{n}\sum_{t=m+1}^{n} M(t,\hat{\boldsymbol{\beta}})M(t-i,\hat{\boldsymbol{\beta}})$$

$$= \frac{1}{n}\sum_{t=m+1}^{n} [\hat{M}(t,\hat{\boldsymbol{\beta}})\{\hat{M}(t-i,\hat{\boldsymbol{\beta}}) - M(t-i,\hat{\boldsymbol{\beta}})\} + \{\hat{M}(t,\hat{\boldsymbol{\beta}}) - M(t,\hat{\boldsymbol{\beta}})\}M(t-i,\hat{\boldsymbol{\beta}})].$$

Since $\hat{\boldsymbol{\beta}}$ is in $\mathcal{B}$, in the same way that (4.5.17) was derived from (4.5.16), we get

$$\hat{\gamma}^{(\hat{M})}(i) = \frac{1}{n}\sum_{t=m+1}^{n} M(t,\hat{\boldsymbol{\beta}})M(t-i,\hat{\boldsymbol{\beta}}) + O(\frac{1}{n}) \quad a.s. . \qquad (4.5.20)$$

Expanding the leading term of (4.5.20) about $\dot{\boldsymbol{\beta}}$, we have

$$\hat{\gamma}^{(\hat{M})}(i) = \hat{\gamma}^{(M)}(i) + \frac{1}{n}\sum_{t=m+1}^{n}\{\frac{\partial M(t,\dot{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}M(t-i) + M(t)\frac{\partial M(t-i,\dot{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\}'(\hat{\boldsymbol{\beta}} - \dot{\boldsymbol{\beta}})$$

$$+ O(\|\hat{\boldsymbol{\beta}} - \dot{\boldsymbol{\beta}}\|^2) + O(1/n) \quad a.s., \qquad (4.5.21)$$

where $O(\|\hat{\beta}-\dot{\beta}\|^2)$ is due to the second derivatives being almost surely bounded. For instance, from (4.5.3a),

$$\sum_{t=m+1}^{n} \frac{\partial M(t,\overline{\beta})}{\partial \phi_k}\frac{\partial M(t-i,\overline{\beta})}{\partial \phi_j} = \sum_{t=m+1}^{n} U(t-k,\overline{\beta})U(t-i-j,\overline{\beta}), \qquad (4.5.22)$$

where $\overline{\beta}$ lies within the join of $\hat{\beta}$ and $\dot{\beta}$, so $\overline{\beta} \in \mathcal{B}$. But

$$U(t,\overline{\beta}) = -\overline{\phi}^{-1}(B)M(t,\overline{\beta}) = -\overline{\theta}^{-1}(B)W(t) = -\overline{\theta}^{-1}(B)\dot{\phi}^{-1}(B)\dot{\theta}(B)M(t),$$

and the boundedness of (4.5.22) follows from Lemma 4.5.1.

In view of (4.5.3), (4.5.2) and Lemma 4.5.1,

$$\begin{aligned} &n^{-1}\textstyle\sum_{t=m+1}^{n}\{\partial M(t,\dot{\beta})/\partial\phi_j\}M(t-i) = -\sigma^2\dot{\phi}^*_{i-j} + O(q_n) \quad a.s. \\ &n^{-1}\textstyle\sum_{t=m+1}^{n}\{\partial M(t,\dot{\beta})/\partial\theta_j\}M(t-i) = \sigma^2\dot{\theta}^*_{i-j} + O(q_n) \quad a.s.\,. \end{aligned} \qquad (4.5.23a)$$

By a similar discussion,

$$n^{-1}\sum_{t=m+1}^{n} M(t)\{\partial M(t-i,\dot{\beta})/\partial\beta_j\} = O(q_n) \quad a.s., \quad j=1,\ldots,p+q. \qquad (4.5.23b)$$

Noticing, from (4.4.32), that $\hat{\beta}-\dot{\beta} = O(q_n)$; (4.5.21) and (4.5.23) give

$$\hat{\gamma}^{(\hat{M})}(i) = \hat{\gamma}^{(M)}(i) - \sigma^2\sum_{j=1}^{p}\dot{\phi}^*_{i-j}(\hat{\phi}_j-\dot{\phi}_j) + \sigma^2\sum_{j=1}^{q}\dot{\theta}^*_{i-j}(\hat{\theta}_j-\dot{\theta}_j) + O(q_n^2) \quad a.s.\,.$$

$$i = 0,1,\ldots,l. \qquad (4.5.24)$$

Then (4.4.35) follows from (4.5.24) and (4.5.13). It is easy to check (4.4.36) from (4.4.33) and (4.4.34).

4.5.2 Proof of (4.4.23)

$$\begin{aligned} &\hat{M}(t+r')\hat{M}(t+s') - M(t+r')M(t+s') \\ &= \{\hat{M}(t+r') - M(t+r')\}M(t+s') + M(t+r')\{\hat{M}(t+s') - M(t+s')\} \\ &\qquad + \{\hat{M}(t+r') - M(t+r')\}\{\hat{M}(t+s') - M(t+s')\}. \end{aligned} \qquad (4.5.25)$$

Using material (4.4.11) through (4.4.15) and (4.5.15)

$$\hat{M}(t) - M(t) = \{M(t,\hat{\beta}) - M(t,\dot{\beta})\} + \{\hat{M}(t,\hat{\beta}) - M(t,\hat{\beta})\}$$

$$= [\hat{\theta}^{-1}(B)\dot{\theta}^{-1}(B)\{\dot{\theta}(B) - \hat{\theta}(B)\}\hat{\phi}(B) + \dot{\theta}^{-1}(B)\{\hat{\phi}(B) - \dot{\phi}(B)\}]W(t)$$

$$-\{\hat{\theta}^{-1}(B)\hat{\phi}(B)\}\tilde{W}(t). \tag{4.5.26}$$

First, due to (4.4.32), $\hat{\boldsymbol{\beta}} - \dot{\boldsymbol{\beta}} = O(q_n)$ a.s.; and, hence, for sufficiently large n, $\hat{\boldsymbol{\beta}}$ falls in $\mathcal{B}$ [the closed neighbourhood of $\dot{\boldsymbol{\beta}}$ throughout which (1.1.3) holds uniformly]. Second, since $\hat{\boldsymbol{\beta}}$ falls in $\mathcal{B}$, the coefficients of $\hat{\theta}^{-1}(B)$ $\dot{\theta}^{-1}(B)\hat{\phi}(B)$, $\dot{\theta}^{-1}(B)$ and $\hat{\theta}^{-1}(B)\hat{\phi}(B)$ are dominated geometrically, analogously to the way ϕ_s^* was in (4.4.45). Third, since $W(t) = o(\mid t \mid^{1/4})$ *a.s.*, and, from (4.5.15), $\tilde{W}(t) = 0$ for $t \geq d$; so, for t in the range $[m+1, n]$, the first term on the RHS of (4.5.26) is $O(q_n \sum_{j=0}^{\infty} e^{-c_1 j} \mid t - j \mid^{1/4})$ $= O(q_n n^{1/4})$ *a.s.*, the second is $O(\sum_{j=t-d}^{\infty} e^{-c_1 j} \mid t - j \mid^{1/4}) = O(e^{-c_1 t}) = O(1)$ *a.s.*, while the third is zero. Thus, in the range $[m+1, n]$,

$$\hat{M}(t) - M(t) = O(1) \quad a.s. \quad . \tag{4.5.27}$$

But in this range, $M(t) = o(t^{1/4}) \leq o(n^{1/4})$ *a.s.*, so (4.4.23) follows from (4.5.25) and (4.5.27).

References

An, H.Z. and Chen, Z.G. (1986). The identification of ARMA processes. *J. Appl. Prob.* **23A**, 75-87.

An, H.Z., Chen, Z.G. and Hannan, E.J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.* **10**, 926-936.

Anderson, O.D. (1979). Formulae for the expected values of the sampled variance and covariances from series generated by autoregressive integrated moving average processes of order (p, d, q). *Sankhyā Ser. B* **41**, 177-195.

Anderson, O.D. (1985). Practical examples of ARUMA modelling. In *Time Series Analysis: Theory and Practice 7.* (Edited by O.D. Anderson), North-Holland, Amsterdam, 15-21.

Anderson, O.D. (1990). Small-sample autocorrelation structure for long-memory time series. *J. Opl Res. Soc.* **41**, 735-754.

Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* John Wiley & Sons, New York.

Beyer, W.H. (1987). *CRC Handbook of Mathematical Sciences* (6th Edition). CRC Press, Boca Raton (FL).

Billingsley, P. (1968). *Convergence of Probability Measure.* John Wiley & Sons, New York.

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control* (Revised Edition). Holden-Day, San Francisco.

Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367-401.

Chen, Z.G. (1988a). Consistent estimates for hidden frequencies in a linear process. *Adv. Appl. Prob.* **20**, 295-314.

Chen, Z.G. (1988b). An alternative consistent procedure for detecting hidden frequencies. *J. Time Ser. Anal.* **9**, 301-317.

Chen, Z.G. (1990). An extension of Lai and Wei's law of the iterated logarithm with applications to time series analysis and regression. *J. Multivariate Anal.* **32**, 55-69.

Cressie, N. (1988). A graphical procedure for determining nonstationarity in time series. *J. Amer. Statist. Assoc.* **83**, 1108-1116. Correction **85**, 270 (1990).

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74**, 427-431.

Doob, J. (1953). *Stochastic Processes.* John Wiley & Sons, New York.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* John Wiley & Sons, New York.

Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Hannan, E.J. (1970). *Multiple Time Series.* John Wiley & Sons, New York.

Hannan, E.J. (1973). Asymptotic theory of linear time-series models. *J. Appl. Prob.* **10**, 130-145.

Hannan, E.J. (1980). The estimation of the order of an ARMA series. *Ann. Statist.* **8**, 1071-1081.

Hannan, E.J. and Kavalieris, L. (1983). The convergence of autocorrelations and autoregressions. *Aust. J. Statist.* **23**, 287-297.

Hannan, E.J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. Biometrika **69**, 81-94.

Hasza, D.P. (1980). The asymptotic distribution of the sample autocorrelation for an integrated ARMA process. *J. Amer. Statist. Assoc.* **75**, 349-352.

Hildebrand, F.B. (1968). *Finite-Difference Equations and Solutions.* Prentice Hall, Englewood Cliffs (NJ).

Huang, D. (1988). Estimating the order and parameters for ARUMA model. Submitted for publication.

Jowett, G.H. (1952). The accuracy of systematic sampling from conveyor belts. *Appl. Statist.* **1**, 50-59.

Kolmogorov, A.N. (1941). The local structure of turbulence in an incompressible fluid at very large Reynolds numbers. *Doklady Akademii Nauk SSSR* **30**, 229-303.

Lakshmikantham, V. and Trigiante, D. (1988). *Theory of Difference Equations: Numerical Methods and Applications.* Academic Press, Boston.

Lai, T.L. and Wei, C.Z. (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.* **13**, 1-23.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246-1266.

Matheron, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Prob.* **5**, 439-468.

McLeod, A.I. (1978). On the distribution of residual autocorrelations in Box-Jenkins models. *J. R. Statist. Soc. Ser. B,* **40**, 296-302.

Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335-346.

Priestley, M.B. (1980). *Spectral Analysis and Time Series.* Vol. I, Academic Press, London.

Said, S.E. and Dickey, D.A. (1985). Hypothesis testing in ARIMA $(p, 1, q)$ models. *J. Amer. Statist. Assoc.* **80**, 369-374.

Stout, W.F. (1970). The Hartman-Wintner law of the iterated logarithm for martingales. *Ann. Math. Statist.*, **41**, 2158-2160.

Tiao, G.C. and Tsay, R.S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Statist.* **11**, 856-871.

Tsay, R.S. and Tiao, G.C. (1984). Consistent estimates of autoregressive parameters and the extended sample autocorrelation function of nonstationary and stationary ARMA models. *J. Amer. Statist. Assoc.* **79**, 84-96.

White, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188-1197.

Wichern, D.W. (1973). The behaviour of the sample autocorrelation function for an integrated moving average process. *Biometrika* **60**, 235-239.

Yaglom, A.M. (1962). *An Introduction to the Theory of Stationary Random Functions.* Prentice-Hall, Englewood Cliffs (NJ).

Yaglom, A.M. (1987). *Correlation Theory of Stationary and Related Random Functions I: Basic Results.* Springer-Verlag, New York.

Yajima, Y. (1985a). Estimation of the degree of differencing of an ARIMA process. *Ann. Inst. Statist. Math. Part A* **37**, 389-408.

Yajima, Y. (1985b). Asymptotic properties of the sample autocorrelations and partial autocorrelations of the multiplicative ARIMA process. *J. Time Ser. Anal.* **6**, 187-201.